Loanwords in Selice Romani, an Indo-Aryan language of Slovakia
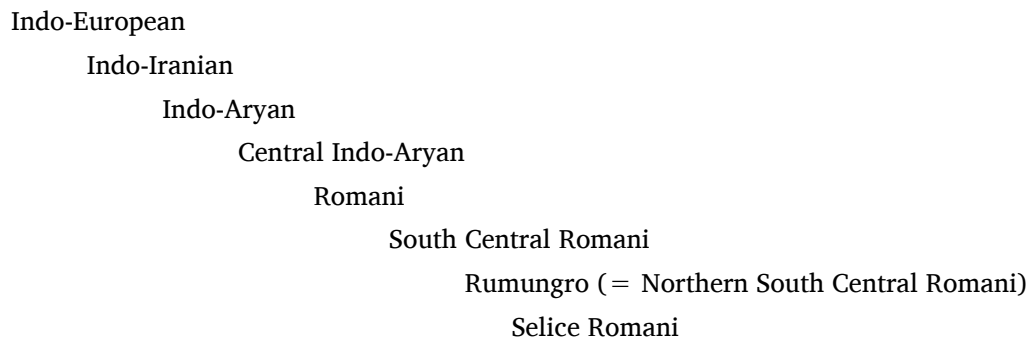
Viktor Elšík

1. The language and its speakers

Romani is an Indo-Aryan (Indo-Iranian, Indo-European) language, whose numerous and rather divergent dialects are spoken by several millions of "Gypsies" – *Roma, Sinti, Mānuš, Kāle* and other related groups – throughout Europe and elsewhere. The variety under description, Selice Romani, is a dialect of Romani spoken by ca. 1,350 Romani inhabitants of the multiethnic village of Selice (Hungarian *Sókszelőce*, Romani *Šóka*) in southwestern Slovakia. Selice Romani is part of a linguistic continuum of closely related Romani dialects spoken in southwestern and south-central Slovakia and in north-central Hungary, which together form the Northern subgroup of the South Central group of Romani dialects (cf. Boretzky 1999; Elšíket al. 1999).[1] The Northern South Central dialects are often refered to as *Rumungro* in Romani linguistics (e.g. Matras 2002) and I will also adopt this term here for its brevity. Although all Rumungro varieties have been influenced by Hungarian, most Rumungro speakers presently live in ethnically Slovak parts of Slovakia and are Slovak bilinguals, whereas an overwhelming majority of Rumungro communities in Hungary and in the Hungarian parts of Slovakia have undergone language shift to Hungarian (cf. Elšík 2003). Selice Romani is one of the few extant Rumungro varieties whose speakers are Hungarian bilinguals.

---

[1] Varieties of the other, Southern (or Vendic), subgroup of the South Central dialects of Romani are spoken in western Hungary, the Austrian Burgenland, and the Slovenian Prekmurje.

The genealogical affiliation of Selice Romani is shown in Figure 1.[2] While I will discuss loanwords into all ancestor varieties of present-day Selice Romani, commencing with Proto-Indo-European, the term *Romani* will only be applied, as is usual, to the part of the variety's genealogical lineage that starts at "the point at which the language became sufficiently distinct from other related *Indo-Aryan* idioms to be classified as an entity in its own right" (Matras 2002: 18; emphasis mine). *Early Romani* is the undocumented, but partly reconstructed, common ancestor of all present-day Romani dialects, which was spoken prior to the dispersion of Romani-speaking groups throughout Europe and the consequent split into dialects (cf. Elšík & Matras 2006: 68–84). *Proto-Romani* (or **Ḍommānī*, cf. Tálos 1999) then covers the pre-Early Romani stages of Romani (but cf. Matras (2002: 18) for a slightly different use of the term). *Pre-split* loanwords are those that can be reconstructed to have been present in Early Romani, while *post-split* loanwords are dialect-specific within Romani. *Pre-Selice Romani* refers to the post-Early Romani ancestor varieties of present-day Selice Romani.

Figure 1: *Genealogical affiliation of Selice Romani*

       Indo-European
           Indo-Iranian
               Indo-Aryan
                    Central Indo-Aryan
                        Romani
                             South Central Romani
                                   Rumungro (= Northern South Central Romani)
                                   Selice Romani

---

[2] Note, however, that the character of Romani dialect groups is a controversial issue: although they may have resulted from separate migrations of Romani speakers out of Asia Minor or the southern Balkans, and so conform well to the family tree model (Boretzky 1999; Boretzky & Igla 2004), they may also have developed *in situ* due to feature diffusion within Romani, and so represent a convenient reference grid rather than genealogical units (Matras 2002, 2005). While I tend to see more evidence for the separate migration scenario in the case of the South Central Romani group (Elšík 2006), the issue certainly requires further research.

Three ethnic groups are represented in the village of Selice:[3] Hungarians, and two distinct Romani groups, viz. the "Hungarian" Roms, most of whom are native speakers of the dialect under description, and the much less numerous "Vlax" Roms, who speak a different Romani dialect natively (see §3.7). Both Romani groups use the plain ethnonym *Rom* for their own group and both are called *cigányok* 'Gypsies' by Hungarians, although the Hungarian villagers clearly differentiate between *magyar cigányok* 'Hungarian Gypsies' and *oláh cigányok* 'Romanian Gypsies', i.e. the Vlax Roms. The former are referred to as *Rumungri* by the latter, who are in turn called *Pojáki* by the former. Until the 1970s, the Hungarian Roms of Selice inhabited a separate, densely inhabited, neighbourhood of one-room adobe houses on the southeastern outskirts of the village. Presently, however, they live in regular houses, interspersed among the Hungarian population. The Vlax Roms have been based in Selice for more than a century, though they were semi-itinerant until 1958, when the Czechoslovak authorities forced them to settle. Their small colony is still located on the northwestern outskirts of Selice. If counted together, the two Romani groups slightly outnumber the Hungarian population of the village.[4] Until recently, however, the Hungarians were in a demographic majority and they remain the economically and politically dominant group in the village.

Selice Romani is prevalently an oral language. Some Hungarian Roms of Selice are able to write letters or text messages in Romani but the language is not used for regular written communication. Nor is it used in mass media or in formal education. Although Romani in general is an officially recognized

---

[3] A score of ethnic Slovaks and Czechs and a couple of Ruthenians and Poles have married into Hungarian or Romani families. The once numerous Hungarian-speaking Jewish community of Selice was completele annihilated during the Holocaust; the single living survivor does not live in the village anymore.

[4] Roms are taken here to be the people who identify themselves as Roms in most informal social contexts and/or who are identified as Roms/Gypsies by other locals. (Most, though not all, Roms thus defined speak Romani natively.) However, only 3% and 4% of the villagers declared Romani ethnicity in the 1991 and 2001 censuses respectively, which amounts to ca. 7% of the Romani population; two thirds of Selice Roms declared Slovak ethnicity and a fifth declared Hungarian ethnicity.

language in Slovakia, there is no recognition of the Rumungro dialect specifically and, so far, there have been no attempts at its standardization. While all Hungarian Roms of Selice born before 1975 or so are native speakers of Selice Romani, in some families children are presently spoken to only in Hungarian and/or Slovak, and left to acquire some competence in Romani in adolescent and adult peer groups, if at all. Thus, Selice Romani is not a safe language, though it is not seriously endangered yet. Interestingly, many Hungarian villagers understand Selice Romani well, although only a few have some active competence in it and I know of no fluent speakers. (See §3.7 for more details on the current contact situation.)

Map 1: *Geographical setting of Selice Romani*

2. Sources of data

All the Selice Romani data in the subdatabase stem from my linguistic fieldwork, which has been carried out during short but numerous fieldtrips to Selice since 1997. I have worked especially with one middle-age female speaker and with people, of both genders and different generations, from within her extended family. Thus, the variety of Selice Romani described here represents a familiolect rather than the local dialect of the Hungarian Roms in general. This is important to stress, as it seems that the Selice Romani lexicon shows significant variation across different groups of speakers, especially with regard to the number of loanwords from Hungarian.[5] In addition to her native language, my main consultant speaks Hungarian, Slovak and Czech fluently, and she has some basic competence in Russian. While a great many of the words in the subdatabase have been acquired through analysis of spontaneous narratives and conversations, all of these have been re-checked with my consultants. A significant part of the subdatabase entries, a third or so, stem from direct lexical elicitation.

Many Early Romani etymologies, including those of pre-split loanwords, have been discussed at least in some of the previous lexical and/or etymological studies on Romani (e.g. Pott 1844–1845, Ascoli 1865, Miklosich 1872–1881, Sampson 1926, Wolf 1960, Valtonen 1972, Vekerdi 1983 [2000], Soravia 1988, Boretzky & Igla 1994, Mānuš 1994, Mānušs et al. 1997, Tálos 1999). Several publications on individual layers of lexical borrowings into Romani are mentioned in §3. I have drawn especially on two sound sources, Boretzky & Igla (1994) (cf. Kostov 1996, Matras 1996) and Mānušs et al. 1997 (cf. Bakker 1999), in etymologizing pre-split loanwords in Selice Romani, while most etymologies of post-split loanwords, including all etymologies of loanwords from Hungarian, Slovak and Czech, are my own. Finally, I have consulted several publications (Beníšek 2006; Buck 1949; Burrow & Emeneau 1960, 1984; Kuiper 1948, 1991; Lubotsky 2001; Mayrhofer 1986–2001; Turner 1962–1966; Witzel 1999a,

---

[5] On the other hand, Selice Romani exhibits a great degree of homogeneity as far as its morphosyntax and phonology is concerned.

1999b, 1999c) in order to identify loanwords into the Old Indo-Aryan and earlier stages of Selice Romani, which, for obvious reasons, have hardly ever been considered in etymological studies on Romani.

3. Contact situations

Selice Romani and its ancestor varieties have come into contact with a number of different languages in a variety of contact situations, including in all likelihood language shift (see §3.2). This section is structured chronologically into periods characterized by contact with a certain language or, more often, with a cluster of languages that may be conveniently discussed together. Although we lack any direct evidence, it is clear that at least after the out-migration of Romani speakers from the Indian subcontinent, the speakers of the immediate contact languages of Romani were overwhelmingly dominant numerically and politically with regard to the Roms. Extrapolating from the similar current demographic and political conditions of Romani in Europe, we may reasonably assume widespread bilingualism among the Roms during their migrations (§3.4–6). As the current contact situation (§3.7) clearly indicates, we must always allow for plurilingualism of the speakers rather than mere bilingualism and for periods of overlap of contact with different languages.

3.1. Contact with non-Indo-European Central Asian languages

Being an Indo-Iranian language, Selice Romani inherits some of the loanwords into Proto-Indo-Iranian that had been acquired before the Aryans arrived in the Indian subcontinent. The source languages of these loanwords remain unidentified, although some authors hypothesize that they mostly represent the non-Indo-European element of ancient Central Asia, specifically the language (or languages) of the Bactria-Margiana Archaeological Complex in the Amu Darya region (e.g. Witzel 1999a: 54; 2003: 52; Lubotsky 2001). While the source forms of the suggested loanwords are unattested, criteria such as irregularity with

regard to the Indo-European phonological, phonotactic and morphological patterns, together with the restricted distribution of the etyma within Indo-European, are used in establishing their loanword status (cf. Lubotsky 2001: 301–305).

Reviewing all Proto-Indo-Iranian words that are unattested elsewhere in Indo-European, Lubotsky (2001) argues that many of them are likely to have been borrowed in Central Asia. Of these probable loanwords, Proto-Indo-Iranian *matsia-* 'fish', * r̥ši-* 'seer', *sūčī-* 'needle', and *u̯r̥tka-* 'kidney' have survived into Selice Romani (see Appendix; note Proto-Indo-Iranian 'seer', 'kidney' > Selice Romani 'priest', 'liver'). In addition, the borrowed Proto-Indo-Iranian *u̯arā́jʰa-* 'wild boar' might be reflected in Selice Romani *bálo* 'pig', if Mānušs et al. (1997: 28) are correct in deriving the Romani word from Old Indo-Aryan *varāhá-* 'wild boar' (cf. Turner 1962–1966: 520 and Boretzky & Igla 1994: 19 for a different view). The Selice Romani verb *khand-* 'to smell' is based on a lost noun (reconstructable for Early Romani) that continued the borrowed Proto-Indo-Iranian noun *gandʰ/t-* 'smell'. A few more of Lubotsky's loanwords have been lost in Selice Romani but are continued in other Romani dialects ('donkey', 'tree', and perhaps also 'well, source'). Of a different origin – perhaps Burushaski, perhaps Semitic, perhaps Anatolian (cf. Mayrhofer 1986–2001: I, 499; Witzel 1999a: 29, 55) – might be the Proto-Indo-Iranian etymon for 'wheat', whose Old Indo-Aryan reflex *godhū́ma-* has developed into Early Romani **giv* (e.g. Turner 1962–1966: 230). The Selice Romani equivalent *šužo jiv* 'wheat', which can be literally translated as 'clean snow', must have developed through confusion of an older **ďiv* 'wheat' (still attested in closely related Rumungro dialects, cf. Vekerdi 2000: 56) and the near-homonymous noun *jiv* 'snow' (which reflects Proto-Indo-European **ǵʰim-* 'cold etc.', e.g. Mayrhofer 1986–2001: II, 815).

Finally, Proto-Indo-European *medʰu-* 'sweet drink, honey' is, according to Witzel (1999a: 55–56), a loanword from an unknown paleo-Eurasian language of eastern Europe or northern Central Asia. If Boretzky & Igla (1994: 183) are correct in deriving Romani *mol* 'wine' from Old Indo-Aryan *mádhu-* 'honey, mead', then this etymon may be the oldest quotable loanword in Selice Romani. However, a much later borrowing into Romani of Persian *mol* 'wine' (e.g. Turner

1962–1966: 562; Mānušs et al. 1997: 87), itself of the same origin, appears to be a more convincing hypothesis on both formal and semantic grounds.

3.2. Contact with non-Indo-Aryan Indian languages

As an Indo-Aryan language, Selice Romani inherits traces of linguistic contacts of its Old and Middle Indo-Aryan ancestor varieties with non-Indo-Aryan languages of India. Kuiper (1991) has shown that already Rgveda, the pre-iron age Old Indo-Aryan text of the Greater Panjab, contains several hundreds of clearly non-Indo-Aryan words. While the presence of Dravidian loanwords in Old Indo-Aryan has long been recognized (e.g. Burrow 1945, 1946, 1947–8; Burrow & Emeneau 1960/1984; Southworth 2005a, 2005b), Witzel (1999a, 1999b) argues that they started to enter the language only in the middle and late Rgvedic periods. The earliest Rgvedic period, on the other hand, is characterized by loanwords from undocumented Greater Panjab substrates. Following Kuiper's (e.g. 1948, 1991) work on Proto-Munda loanwords in Old Indo-Aryan, Witzel (1999a) refers to the major Rgvedic substrate as *Para-Mundic* and considers it to be a western variety of Austroasiatic. The number of both Dravidian and (Para/Proto-)Munda loanwords in Indo-Aryan increases in post-Vedic times (Burrow 1973: 386, Witzel 1999a: 34). In addition, a number of unidentified substrate languages, such as Masica's (1979) Gangetic *Language X*, have been suggested to have contributed loanwords to regional varieties of Indo-Aryan.

Selice Romani retains over a dozen of non-Indo-Aryan Indian loanwords into Indo-Aryan, which are, with a few exceptions (e.g. 'sack' or 'straw'), represented in the Loanword Typology (LWT) meaning list. The bulk of the loanwords are attested in, or have been reconstructed for, Old Indo-Aryan, though a few may be of a later or local origin. For example, Romani *purum* 'onion', a possible loanword from Dravidian (cf. Tamil *pūṇḍu* 'onion, garlic', Mānuš 1994: 34;

Mānušs et al. 1997: 106), appears to be isolated within Indo-Aryan.[6] Some of the Indian loanwords in Romani have a more or less established Dravidian etymology (Burrow & Emeneau 1960/1984; Turner 1962–1966), while others continue probable or possible loanwords from Proto-Munda (Kuiper 1948). It is possible that the Romani word *murš* 'man, male' continues a loanword of Proto-Burushaski *mruža/mruša-* 'Burusho' into Old Indo-Aryan.[7]

Certainly the most telling Indian loanword in Romani is the ethnic autonym of Roms, cf. Early Romani *ṛom* *'Rom; Romani married man; Romani husband'.[8] Its ancestor form, Old Indo-Aryan *ḍōmba-*, which also survives as the name of other Indian-origin ethnic groups in the Middle East and of various low castes in northern India (cf. Briggs 1953), is clearly of Munda provenance (Kuiper 1948: 87; Turner 1962–1966: 313; Beníšek 2006). This indicates (though does not prove) that the Ḍōmba were originally a Munda-speaking group who shifted to an Indo-Aryan language (Vekerdi 1981; Beníšek 2006). On account of the late attestation of the term *ḍōmba-* in Indo-Aryan, viz. in the sixth century CE, Beníšek (2006: 23–24) suggests that the shift did not take place before the beginning of the Common Era.

---

[6] It certainly does not continue Old Indo-Aryan *palāṇḍu-* 'onion', of unclear etymology (Mayrhofer 1996: II, 102) and probably also a borrowing, on account of the "suspicious" cluster /ṇḍ/ (cf. Witzel 1999a: 11, 43).

[7] Traditionally, the Romani word has been explained as a contamination of Old Indo-Aryan *manuṣyà-* 'human being', which itself results in Romani *manuš*, with Old Indo-Aryan *puruṣa-* 'man' (e.g. Turner 1962–1966: 564). The latter has been suggested to be based on the Proto-Burushaski form (Witzel 1999c) but given the presence of *m*-initial forms such as Multani and Parya *muṛs*, Sindhi *mursu* etc. in the Indian North West, we may perhaps derive the Romani word directly from an unattested *m*-initial Old Indo-Aryan form.

[8] While some groups of Romani speakers have replaced this original ethnonym by various innovative autonyms (e.g. Matras 1999, 2002), all Romani dialects retain the word's secondary meaning '(Rom) husband', whose development has been elucidated by Beníšek (2006: 14–17). In some dialects, the word can only be used to refer to husbands of the Romani ethnicity in its secondary meaning, while in others, including Selice Romani, it has acquired an ethnically neutral meaning 'husband'.

3.3. Contact with other Indo-Aryan languages

It is likely that, in addition to borrowing from the non-Indo-Aryan Indian languages, there was also lexical borrowing from other Indo-Aryan varieties into the Indo-Aryan ancestor varieties of Romani. First, there may have been loanwords into Proto-Romani from literary Indo-Aryan languages, though – assuming that Proto-Romani did not have any literate speakers – they would have had to be acquired through mediation of other vernaculars. For example, Turner (1926: 151) suggests that Romani *truš* 'thirst' and *rašaj* 'priest', both retained in Selice Romani, may reflect early loanwords from Sanskrit. In a later publication he only derives the latter from an unattested North Western Prakrit form (Turner 1962–1966: 118), which brings us to a second, geographical, point:

Turner (1926) argues convincingly that Proto-Romani originated as a Central Indo-Aryan variety and, somewhat less convincingly (cf. Woolner 1928; Beníšek 2006: 23–24), that it must have severed its connection with the Central group before the third century BCE. He also claims that Proto-Romani speakers then migrated to the Indian northwest, which was actually long (e.g. still in Turner 1924: 41) believed to be the original home of Proto-Romani; there they spent several centuries, borrowing words, including several that can be identified specifically as Nortwestern Indo-Aryan or even "Dardic." The ones Turner (1926: 156, 174) explicitly mentions are reflected in Selice Romani as *štár* 'four', *šó* 'six' and *murš* 'man, male'. However, as Matras (2002: 47) points out, the lexical evidence for the Northwestern contact of Proto-Romani is "marginal and largely inconclusive." Indeed, Turner (1962–1966: 742–743) himself appears to have later revised his Dardic hypothesis regarding the origin of the Romani numeral 'six', deriving it instead from a separate Old Indo-Aryan form, and he no more mentions the possible Dardic origin of the other Romani forms.

3.4. Contact with Middle-Eastern languages

While hypotheses about the time of the out-migration of Proto-Romani speakers from India vary tremendously, ranging between the fourth century BCE and the

eleventh century CE, Matras' (2002: 18) suggestion that the ancestors of the Roms left the subcontinent some time in the eighth or ninth century CE cannot be wildly off the mark. Between this period and the arrival of the Roms in the Byzantine Empire (see §3.5), Proto-Romani was in contact with several Middle Eastern languages, as evidenced by loanwords attested in various Romani dialects and hence reconstructable for Early Romani:

First, there are a relatively high number of **Iranian** loanwords in Romani. Boretzky & Igla (1994: 329–331) list 67 possible Iranianisms, of which over three dozen are quite certain, while Hancock (1995) includes as many as 119 potential loanwords from Iranian, though many of these are obviously recent, dialect-specific, borrowings into Romani dialects of the Balkans via Turkish and other Balkan languages (cf. Matras 2002: 23). Additional lexical Iranianisms not identified or classified as such in either of the above lists are identified especially in Mānušs et al. (1997). The overwhelming majority of Iranian loanwords in Romani can be derived from (late) Middle Persian, although many allow for, and some appear to require, a different source. Kurdish and Ossetic are widely held to have contributed a few loanwords each, e.g. Early Romani *kirivó* 'godfather' < Kurdish *kirîv* (Mānušs *et al.* 1997: 72) and Early Romani *vṛdón* 'cart, wagon' < Ossetic *wardon* (e.g. Boretzky & Igla 1994: 301, 331; but cf. also Middle Persian *wardyūn*). Selice Romani retains two dozen Iranian loanwords from the larger Early Romani pool, including *zijand* 'damage, pity' from Persian *ziyān* 'damage [etc.]' (my etymology).[9] Most of the Iranian loanwords in Selice Romani are represented in the LWT meaning list, with the exception of a possibility particle and nouns meaning 'strength, force, power', 'whip', and 'co-father-in-law'.

Second, the Romani lexicon contains loanwords from **Armenian** (many of which are themselves loanwords from Iranian, and sometimes difficult to distinguish from immediate Iranianisms). Their number is somewhat lower than

---

[9] The form of the noun *ziján-i* 'damage' in some Romani dialects of the Balkans (e.g. in the South Vlax dialect of Ajia Varvara, Athens; cf. Messing 1988: 140, Friedman 1989) clearly indicates that it is a relatively recent Turkism (of Persian origin). On the other hand, the form of the Selice Romani word makes it clear that it continues a Proto-Romani loanword from Persian.

that of Iranian loanwords, though still relatively important: recent overviews list 34 (Hancock 1987), 41 (Boretzky & Igla 1994: 331–332), or 51 (Boretzky 1995) possible items, of which around two dozens are quite certain (cf. Matras 2002: 23). Selice Romani retains only nine certain or probable loanwords from Armenian, one of which is not represented in the LWT meaning list: *pativ-ake* 'in vain, for free', an adverbialized dative of the noun **pativ* 'honour' < Armenian *patiw*, which has been lost in the variety.

Finally, four Romani nouns have been suggested to be loanwords from **Georgian**: 'plum', 'suet, tallow' (e.g. Pobożniak 1964: 79), 'eyelash' (Friedman 1988), and 'sand' (Grant 2003: 27). None of these etyma have survived into Selice Romani: they have been replaced either by more recent loanwords or through a dialect-specific semantic shift of an indigenous word (viz. 'sand' < 'dust, powder').

Since "[a] thorough investigation of the Iranian element in Romani from an Iranist's point of view is still missing" (Matras 2002: 23), we cannot exclude that Proto-Romani was also in contact with other Iranian languages than those mentioned above. If, however, the lack of loanwords from East Iranian languages (with the exception of Ossetic, spoken in the Caucasus) and Balochi turns out to be genuine, we may hypothesize a relatively rapid migration of the ancestors of the Roms out of the Indian subcontinent to Khorasan, a more likely place, it appears, for their acquision of Persian loanwords than Fars. The further migration route is likewise far from certain: Boretzky (1995) considers the possibility that the few Georgian words in Romani were borrowed via Armenian. Matras (2002: 25), in a similar vain, suggests that both the Georgian and the Ossetic loanwords may have been transmitted via other sources. Also, most if not all of the suggested Ossetic loanwords allow for alternative, Iranian or Armenian, etymologies. Considering all this plus the well-known fact that Armenian was also spoken in eastern Anatolia, it is quite possible that Proto-Romani speakers never actually inhabited the southern Caucasus. Indeed, Matras (1996, 2002: 25) suggests that the contact of Romani with Armenian and Western Iranian could have taken place simultaneously with its contact with Byzantine Greek. This is compatible with, though not implied by, Toropov's

(2004: 15) convincing argument that Romani contact with Armenian must have occurred by the ninth century CE.

Important for the reconstruction of Romani migrations is the lack of any unambiguous pre-split loanwords from Turkic, whether immediate or mediated by other languages. Ultimate Arabisms are very rare and most likely mediated by other Middle Eastern languages (Matras 2002: 25). Selice Romani retains a single Arabism, viz. *humer* 'boiled or baked dough; pastry; noodles', which has been borrowed into Romani via Persian and/or Armenian. Interestingly, Berger (1959) suggests a number of Burushaski etymologies for Romani, which however are mostly rejected as unconvincing by Matras (2002: 24). One of Berger's Burushaskisms, reflected in Selice Romani as *cid-* 'to pull; draw; suck', is deemed possible by Matras but it receives a more convincing Indo-Aryan etymology in Tálos (1999: 257), and so we may actually dispense with the assumption of the presence of Romani speakers in the Karakoram Mountains on their way out of India.

3.5. Contact with Greek

While the first historical records of the presence of Gypsies in the Byzantine Empire originate from the late eleventh century CE (e.g. Soulis 1961), Tzitzilis (2001: 327–8) argues on linguistic grounds that Romani contact with Greek must have occurred by the tenth century. He also suggests that the oldest layer of Hellenisms in Romani are loanwords from Pontic and Cappadocian dialects of Medieval Greek, which of course also makes sense geographically. Differing degrees of morphological integration of Greek loanwords may reflect different layers of contact (see §5.2). For example, Greek *ðróm-os* 'way' is fully integrated as *drom* in Romani, and is likely to be an earlier loanword than that of Greek *fór-os* 'square; market', which retains its Greek nominative inflections in Romani. The fact that Greek is the source of numerous inflectional and derivational affixes in Romani (e.g. Boretzky & Igla 1991, Bakker 1997) and the model of radical morphosyntactic Balkanization-cum-Hellenization of the language (e.g. Friedman 1986, 2000; Matras 1994, 1995) suggests that contact with Greek

involved fluent bilingualism of adult Romani speakers. Since most of the Greek-origin grammatical component is shared by all present-day Romani dialects, we may safely assume a relatively homogeneous speech community at the time of (early) Greek contact and locate Early Romani, the common ancestor of all modern Romani dialects, in the Byzantine period.

Selice Romani retains three dozen Greek loanwords, a third of which are not represented in the LWT-based subdatabase, including nouns meaning 'cabbage', 'carrot', 'fairy tale', 'lap', 'jelly', and several function words. This number contrasts, for example, with twice as high a number of Hellenisms in a familiolect of Welsh Romani (Sampson 1926, counted in Grant 2003: 29).[10] Both numbers certainly represent a mere fraction of all Greek loanwords that were in use in Romani during its Byzantine period, as indicated by the sum of Hellenisms that have been retained at least in some modern dialects of Romani outside of the Greek-speaking area. For example, Boretzky & Igla's (1994) dictionary contains a list of 238 loanwords from Greek; Grant (2003) lists over 300 items, of which 260 he considers to be assured or likely; and there are several additional Greek items in Vekerdi (1983 [2000]) and Tzitzilis (2001) not discussed in either of the above. Two loanwords retained in Selice Romani have not been previously identified as Hellenisms, viz. the ethnonyms *ungro* 'Hungarian' and *servo* 'Slovak' < Greek *úngros* and *sérvos* 'Serb', respectively.

3.6. Contact with South Slavic languages

The first historical records of the presence of Gypsies in the South Slavic area originate from the second half of the fourteenth century CE (e.g. Fraser 1992), just before the Ottoman conquest of Bulgaria and Serbia, though the first contacts of Romani speakers with South Slavic are likely to have occurred somewhat earlier. Since early historical records do not discriminate between different Romani groups, we are not in position to date with any precision the

---

[10] Grant (2003: 29) also counts Greek loanwords in other Romani dialects such as Lovari (Vekerdi 1983), but these represent dialect clusters rather than individual local varieties, and so these counts are, strictly speaking, not comparable to the number of loanwords in Selice Romani.

beginning of the South Slavic bilingualism of the ancestors of Selice Romani speakers on historical grounds.

The South Slavic languages contribute almost three dozen loanwords to the subdatabase, which amount to two thirds of all South Slavic loanwords attested in Selice Romani. Those that are not represented in the sample include an ethnic noun refering to non-Roms, which has the source meaning '(the) coarse (one)'; the comparative adjective 'worse', whose suppletive positive-degree counterpart is also a South Slavic loanword; and more. The number of South Slavic loanwords was certainly much higher during the time of South Slavic bilingualism of pre-Selice Romani speakers. In fact, closely related Rumungro varieties retain a number of Slavicisms that have been replaced by Hungarian loanwords in Selice Romani, e.g. 'world', 'foreign', 'to write', and more.

A few South Slavic loanwords have a relatively wide distribution within Romani and may be assumed to have been borrowed into the language before the out-migration of different Romani groups from the southern Balkans and their geographical dispersal throughout Europe (cf. Boretzky & Igla 2004: 9; Boretzky n.d.). One example of such a word is Selice Romani *vodro* 'bed' (cf. Old Church Slavonic *odrŭ* 'bed'), which is also attested, for example, in Welsh and Finnish Romani. Its meaning, too, shows that it must be a relatively old borrowing: the word has undergone various semantic specializations in modern South Slavic languages, e.g. Bulgarian *odăr* 'plank bed', Serbo-Croatian *odar* 'hearse, catafalque', or Slovene *oder* 'platform, plank stand'. Nevertheless, the majority of South Slavic loanwords in Selice Romani are dialect-specific loanwords, most of which are restricted within Romani to the South Central dialect group.

Several South Slavic loanwords in Selice Romani could have originated in any South Slavic idiom, e.g. *zelen-o* 'green' < *zelen*. Mostly, however, the distribution of the source word is restricted within the South Slavic area, and it is often possible to identify the source language quite specifically, due to form and/or meaning peculiarities of the Selice Romani loanword. For example, Selice Romani *erďavo* 'bad, evil, wrong' clearly derives from Serbo-Croatian *rđav* 'rusty; bad, evil', since the other South Slavic languages exhibit very different forms and have not developed the relevant secondary meaning 'bad, evil' (cf. Bulgarian

răždiv, Macedonian *'rǵosan,* Slovene *rjast* 'rusty'). A few Selice Romani words, both within and without the sample, can be identified even more specifically as loanwords from an Ikavian dialect of Serbo-Croatian (Elšík et al. 1999), e.g. *cilo* 'whole; all' < *cio ~ cil-, ninco* 'German' < *nimac ~ nimc-.* While quite a few South Slavic loanwords in Selice Romani must originate in Serbo-Croatian, almost all of them *can,* and so it may well be that Selice Romani acquired almost all of its South Slavic loanwords from a single source.

Although there is no historical documentation of the out-migration of the ancestors of Selice Romani speakers out of the South Slavic linguistic area, it is quite likely that it was part of wider population movements triggered by the Ottoman expansion in the Balkans and towards Hungary and Hapsburg Austria. It is tempting to connect the current presence of the South Central Romani speakers in the western part of historical Hungary to the large-scale re-settlement of Croats to Burgenland (*Gradišće*) and the neighbouring parts of Hungary, including the southwest of present-day Slovakia, which took place especially during the sixteenth century.[11] However, a small piece of linguistic evidence appears to indicate a somewhat later out-migration. The only Turkism among the South Slavic loanwords in pre-Selice Romani, viz. *duhano* 'tobacco' < Serbo-Croatian *duhan* (< Turkish *duhan* 'smoke' < Arabic *duhān*; cf. Buck 1949: 534), denotes a New World plant that was introduced into the Balkans by the Ottomans at the very beginning of the seventeenth century (e.g. Mijatović 2006). This requires that there still was contact between pre-Selice Romani and (the Turkish-influenced varieties of) Serbo-Croatian at this time.[12]

---

[11] For example, the village of Hrvatski Grob, located several dozen kilometers to the northwest of Selice, was founded in 1552 by settlers from the Moslavian region in Croatia. The local Croatian dialect, still spoken by some elders, contains an Ikavian element.

[12] The etymon is also found in Hungarian as *dohány* 'tobacco' and it cannot be excluded that the immediate source of the Selice Romani word is an unattested dialectal Hungarian form **duhan.*

3.7. The current contact situation

All school-age or older native speakers of Selice Romani are plurilingual, speaking two or more languages fluently, in addition to Romani. First of all, they are all fluent and highly competent in **Hungarian**, which they use especially in their everyday communication with the Hungarian villagers but also with those Hungarian Roms of the village and the region who are less competent in Romani or who do not speak or understand Romani at all. Some young children may be monolingual in Romani, although early acquisition of Hungarian appears to be the prevailing pattern nowadays. We do not know when the contact with Hungarian started, neither is it clear when the ancestors of the Hungarian Roms of Selice settled in the village. They retain no memory of their previous homes or migrations and the locals claim that the recently abandoned settlement of the Hungarian Roms (see §1), by far the largest Romani settlement in the region, had been there "from times immemorial." The bilingualism of Selice Romani speakers in Hungarian has certainly lasted for many generations, and quite likely for several centuries.

An overwhelming majority of Hungarian Roms of Selice are also fluent in **Slovak**, which they use especially at schools and outside of the village.[13] Although few ethnic Slovaks live in Selice, Slovak-speaking villages are located nearby, and so it is likely that the first contacts of Selice Romani with Slovak predate the creation of Czechoslovakia in 1918, whereafter Slovak became the official and dominant language of Slovakia. The contact with vernacular Slovak of the region is confirmed by dialectal features in the Slovak of elder Roms and by the form of some established Slovak loanwords in Selice Romani, e.g. *škráteko* 'elf' from Slovak dialectal *škrátek* (cf. standard *škriatok*).[14] Nevertheless, it has been the recent influence of Slovak mass media and schooling that contributed

---

[13] In contrast, some local Hungarians are still monolingual in Hungarian and hardly understand Slovak.

[14] An early contact with Slovak is, incidentally, also suggested by a peculiar semantic shift in the loanword of the Greek ethnonym *sérvos* 'Serb': the fact that Selice Romani *servo* means 'Slovak' appears to indicate that the ancestors of the Hungarian Roms still spoke, or at least understood, South Slavic when they first encountered the Slavic-speaking Slovaks.

to the general Slovak bilingualism among the Hungarian Roms of Selice. Most Hungarian Roms of Selice have also acquired at least passive competence in **Czech** through their exposure to Czech mass media and especially during their employment-related stays in the Czech part of the former Czechoslovakia, where most families spent between ten to thirty years in 1960–1980s. Many Selice Romani speakers, including my main consultant, attended Czech primary schools.

Active competence in other languages is individual and usually acquired during job-related stays in foreign countries. My primary consultant and several members of her family spent a year in Kazakhstan in early 1990s, where they spoke Russian with the locals. I am aware of a single word of Russian origin in Selice Romani, viz. *ďengi* 'money' < *d'en'g'i*, which is a rarely used slang alternative to an indigenous Romani word.

Finally, a few words about the social and linguistic relations between the Hungarian Roms and the Vlax Roms of Selice are in order. Both groups consider their own group to be superior.[15] There is no intermarriage between members of the two groups, and social contact is mostly restricted to economic exchange. The native language of the Vlax Roms is a Lovari-type North **Vlax** dialect of Romani (cf. Boretzky 2003), which is quite different from Selice Romani. In fact, the Hungarian Roms claim that they do not understand much of the dialect of the Vlaxs, and my field observations appear to confirm this. Yet, many Hungarian Roms are aware of certain salient lexical differences between the dialects and take some pride or amusement in citing "typical Vlax words," e.g. *khanči* 'nothing' (cf. Selice Romani *ništa*). All adult Vlax Roms, on the other hand, regularly use Selice Romani, or rather a distinct ethnolect of it, in communication with the Hungarian Roms. Given the mutual disdain, this asymmetrical pattern clearly reflects the demographic asymmetry between the two Romani groups in Selice.

---

[15] To wit: the Hungarian Roms consider themselves to be more civilized and progressive, resenting the wildness and backwardness of the Vlaxs, while the Vlax Roms consider themselves to be the only real and pure Roms, disdaining the Hungarian Roms as assimilated half-Hungarians (hence also the ethnic exonym *Rumungro*, originally **Rom-Ungro* 'Gypsy-Hungarian').

The lack of any significant competence of the Hungarian Roms of Selice in the Vlax dialect makes it unsurprising that there are very few Vlax loanwords in Selice Romani. One of them is *krísa*, a loanword of Vlax *krísi* 'judgement, trial, tribunal, court', itself a loanword from Greek, which is used to refer to a community-internal judicial institution among the Vlaxs (no such institution exists among the Hungarian Roms). The Greek loanword is likely to have been present in Early Romani, then lost in the ancestor variety of Selice Romani, and then – as its meaning and form clearly show – borrowed "again" as a cultural insertion from Vlax.

4. Numbers and kinds of loanwords

4.1. A note on what counts as a loanword

There are 1430 lexemes in the Selice Romani subdatabase, of which 62.6% I classify as loanwords. In the overwhelming majority of instances, the lexemes considered to be loanwords here have been borrowed without any doubt, while a tiny minority of them are merely probable loanwords. In addition, a couple of dozen further words have been suggested to be loanwords (and indeed may be ones), but are not counted as such in this paper, because I do not consider their borrowing etymologies to be fully convincing. In addition to loanwords proper, there are ca. 6% of lexemes in the sample that are merely "created on loan basis" and not counted as loanwords: these are either lexicalized collocations or compounds containing a clear or probable loanword, or (synchronic or merely etymological) derivations from a clear or probable loanword.[16] Semicalques,

---

[16] The Selice Romani noun *žuto* 'yolk', for example, has developed through onomasiological conversion of the adjective *žuto* 'yellow', which is a clear loanword of Serbo-Croatian *žut* 'yellow'. The conversion may have occurred due to pattern borrowing from Hungarian, cf. *sárga* 'yellow' and *(tojás-)sárgá-ja* [(egg-)yellow-3SG.POSS] 'yolk'. Although the (base) form of the Selice Romani noun is identical to that of the borrowed adjective and although the noun's development through conversion may have been contact-induced, the noun is not considered to

which involve borrowing of matter but not borrowing of the whole form of the lexeme, e.g. Selice Romani *vala-kana* vs. Hungarian *vala-mikor* [some-when] 'sometimes', are not considered to be loanwords either. This rather restrictive approach to what counts as a loanword means that the number of words that consist exclusively of indigenous morphemes is significantly smaller than the number of words that are classified as non-loanwords.

4.2. Loanwords by source language

It is often difficult to identify the immediate source language of a loanword precisely, especially due to genealogical relatedness or contact between source languages. For example, Selice Romani *kopaj* 'stick; club' can be a loanword from Pontic Greek, but also from Armenian or Kurdish, which borrowed the Greek word (cf. Tzitzilis 2001: 332). Given this, I find it useful to simplify the quantitative presentation of the data by lumping, in the following cases, several source languages into "contact clusters:" the **Indian** cluster consists of loanwords into Old and Middle Indo-Aryan from (Para/Proto-)Munda and/or Dravidian (see §3.2); the **South Slavic** cluster subsumes any South Slavic source (see §3.6); and, finally, the **Slovak/Czech** cluster consists of loanwords from both Slovak and Czech. In addition, I took a few arbitrary decisions, including the following: loanwords that can originate in Hungarian are counted as Hungarian, even if they can also originate in Slovak/Czech and/or South Slavic; and loanwords that can originate in South Slavic and Slovak/Czech are counted as South Slavic. Table 1 shows the breakdown of sample loanwords by donor language or donor language group:

---

be a loanword, since there is no noun of the relevant form and meaning in the source language (cf. Serbo-Croatian *žumance, žumanjak, žutanjak, žutac* etc. 'yolk').

Table 1: *Loanwords in Selice Romani by source language*

Source language	#	% of words	% of loanwords
pre-Indian	3	0.2	0.3
Indian	12	0.8	1.3
Persian	18	1.3	2.0
Kurdish	1	0.1	0.1
Ossetic	2	0.1	0.2
Armenian	9	0.6	1.0
Greek	25	1.7	2.8
South Slavic	32	2.2	3.6
Hungarian	753.5	52.7	84.2
Slovak/Czech	38	2.7	4.2
Vlax Romani	2	0.1	0.2
Total loanwords	*895.5*	*62.6*	*100.0*
Total words	*1430*	*100.0*	–

Table 1:

Hungarian, the primary current contact language of Selice Romani, is far and away the most important source of loanwords, contributing the bulk of all loanwords and over half of all words in the sample. This statement remains true even if items that may but need not be immediate loanwords from Hungarian are discounted. In addition, there are hundreds of established loanwords from Hungarian that are regularly used in Selice Romani but whose meanings are not represented in the sample. Unsuprisingly, Hungarian is also a frequent source of nonce loanwords in Selice Romani discourse. In contrast, the other contact languages or clusters, including all past contact languages, each contribute less than a twentieth of all loanwords. Although nonce loanwords from Slovak and Czech often occur in the speech of many Selice Romani speakers, the number of established Slovak or Czech loanwords cannot be much higher than the one indicated by the sample. Considering the fact that Selice Romani speakers are fluent active bilinguals in Slovak, and many of them in Czech as well, the great quantitative disproportion between the Hungarian and the Slovak(/Czech) lexical components in Selice Romani is striking. Assuming that the length of

contact is hardly the only factor, the disproportion is in need of a detailed sociolinguistic explanation.

Since there is no space here to discuss in any detail the ultimate and intermediate sources of Selice Romani loanwords, I will restrict myself to a few remarks: The current contact languages Hungarian, Slovak and Czech have mediated a number of loanwords from German, Latin, French, Italian, and other languages. Hungarian is also the immediate source of a number of Slavisms (including recent Slovakisms in the local Hungarian dialect) and Turkisms (mostly of Oghuric affiliation). In addition to direct loanwords from Greek there are also several ultimate Hellenisms in Selice Romani that entered the language via Hungarian, Slovak/Czech or Vlax Romani. On the other hand, immediate contact with Greek also introduced a couple of Latin and ultimately Germanic (via Italian: 'soap') and Turkic (via Slavic: 'Hungarian') words. Direct loanwords from Iranian languages contrast with Iranianisms acquired via Armenian, Hungarian (e.g. 'thousand') or via Turkish and South Slavic ('cotton'). Names of several plants and products originating in South Asia have been re-introduced via European languages (e.g. 'black pepper', 'rice', or 'sugar'). Lexical borrowing has resulted in several etymological doublets in Selice Romani.

4.3. Loanwords by word class

The standard breakdown of sample loanwords by semantic word class is shown in Table 2.[17]

---

[17] The Selice Romani morphosyntactic word classes Verb, Noun, and Adjective closely match the semantic word classes. Almost any individual LWT meaning of a certain semantic word class (as indicated in the database template) can be, provided it is lexicalized at all in Selice Romani, rendered by an expression of the corresponding language-specific morphosyntactic word class. There are only very few exceptions: for example, there is no adjective meaning 'stinking', only a verb meaning 'to stink' in Selice Romani. Consequently, the breakdown of loanwords by Selice Romani word classes would show numbers almost identical to those of Table 2.

Table 2: *Loanwords in Selice Romani by semantic word class (percentages)*

	Hungarian	Czech/Slovak	South Slavic	Greek	Iranian	Indian	Armenian	Vlax Romani	South Slavic	Unidentified	Total loanwords	Non-loanwords
Nouns	59.9	7.2	2.4	2	1.9	0.7	1	0.1	0.1	0.4	75.6	24.4
Verbs	41.2	-	1.2	0.9	1.5	0.3	-	-	-	-	45.1	54.9
Adjectives	42.1	-	4	-	1.6	3.2	-	0.8	-	-	51.7	48.3
Adverbs	50	-	-	-	-	-	-	-	-	-	50	50
Function words	21	0.6	4.2	4.2	0.8	-	-	-	-	-	30.9	69.1
all words	**50.8**	**4.3**	**2.4**	**1.7**	**1.7**	**0.8**	**0.6**	**0.1**	**0.1**	**0.2**	**62.7**	**37.3**

Of all word classes, nouns exhibit the highest proportion of loanwords: over three quarters. The other content word classes lag behind nouns and are roughly similar to one another with regard to loanword proportions: loanwords represent half of all adverbs, just over half of all adjectives, and somewhat less than half of all verbs. However, adverbs only amount to 4 items in the LWT meaning list, and so the proportion of loan-adverbs is clearly beyond statistical significance. In fact, *all* Selice Romani manner adverbs that semantically correspond to Hungarian-origin adjectives are themselves lexical borrowings from Hungarian, rather then internal derivations from the borrowed adjectives, and so the proportion of loan-adverbs could be very different in an extended meaning sample. Finally, function words show the lowest proportion of loanwords: just below a third.

Table 3 displays the proportions of selected diachronic layers of loanwords to all loanwords by word class (the word classes are arranged by decreasing loanword proportions), plus arithmetical differences from the total proportion of this kind. The diachronic layers considered are: loanwords from Hungarian; loanwords from all current contact languages, i.e. Hungarian, Slovak, Czech, and Vlax Romani; and loanwords acquired since the contact with Greek, including those from the current contact languages, i.e. roughly during the last millenium.

Table 3: *Loanwords in Selice Romani by semantic word class and diachronic layer (percentages)*

Word class	Loans	Hungarian		Current L2s		Last 1000 years	
Nouns	75.6	83.3	–0.9	89.3	+0.6	94.9	–0.1
Adjectives	51.7	81.4	–2.8	82.9	–5.8	90.7	–4.3
Adverbs	50.0	100.0	+15.8	100.0	+11.3	100.0	+5.0
Verbs	45.1	91.9	+7.7	91.9	+3.2	96.6	+1.6
Function words	30.9	71.8	–12.4	71.8	–16.9	97.1	+2.1
Total	*62.7*	*84.2*	*0.0*	*88.7*	*0.0*	*95.0*	*0.0*

Hungarian loanwords (and the current loanwords in general) represent over four fifths of all loanwords in any content word class; the proportion is somewhat lower in function words. At least 90% of loanwords of any word class have been borrowed within the last millenium of the history of Selice Romani. The following may also be read off Tables 2 and 3: Hungarian is unique among the source languages in contributing a higher proportion of loan-verbs than that of loan-nouns (with regard to all loanwords of the respective word class). Slovak and Czech only contribute nouns, not other word classes. The LWT meaning list appears to be representative in this respect: although there is an established mechanism for morphological integration of Slovak and Czech verbs (see §5.4), they appear to be overwhelmingly, if not exclusively, nonce loanwords; and there are no established mechanisms for morphological integration of Slovak and Czech adjectives.

4.4. Loanwords by semantic field

The standard breakdown of loanwords by semantic fields is shown in Table 4.

Table 4: *Loanwords in Selice Romani by semantic field (percentages)*

		Hungarian	Czech/Slovak	South Slavic	Greek	Iranian	Indian	Armenian	Vlax Romani	South Slavic	Unidentified	Total loanwords	Non-loanwords
1	The physical world	66.4	0.8	2.4	-	3.2	1.6	-	-	-	-	74.3	25.7
2	Kinship	25.9	1.2	1.2	1.2	-	2.5	-	-	-	-	32.1	67.9
3	Animals	61.1	11.7	1.2	1.2	1.2	-	1.2	-	-	-	77.8	22.2
4	The body	45.9	2	1.4	1.4	3.7	0.9	1.2	-	-	0.6	57.2	42.8
5	Food and drink	44.8	4.1	2.2	1.3	3.2	2.5	0.6	-	-	-	58.8	41.2
6	Clothing and grooming	67	6.4	1.2	2.2	5	1.7	1.7	-	-	1.7	86.7	13.3
7	The house	66.9	15.7	4.8	3.5	-	-	-	-	1.7	-	92.7	7.3
8	Agriculture and vegetation	67.6	10.7	7.5	2.3	-	-	1.7	-	-	-	89.8	10.2
9	Basic actions and technology	52.5	3.3	2	1.3	1.3	-	-	-	-	-	60.5	39.5
10	Motion	48.8	-	2.5	2.5	3.8	-	-	-	-	-	57.6	42.4
11	Possession	39.7	4.3	-	2.1	2.9	-	-	2.1	-	-	51.2	48.8
12	Spatial relations	39.3	2.6	2.6	-	1.3	2	-	-	-	-	47.7	52.3
13	Quantity	19.1	-	5.1	10.2	-	-	2.6	-	-	-	37	63
14	Time	51.3	1.8	0.8	5.4	-	-	-	-	-	-	59.2	40.8
15	Sense perception	46.5	-	4.4	-	-	4.4	-	-	-	-	55.4	44.6
16	Emotions and values	40.3	-	4	4	3.4	-	-	-	-	-	51.7	48.3
17	Cognition	50.7	-	-	-	-	-	-	-	-	-	50.7	49.3
18	Speech and language	56.4	0.8	5	-	-	-	-	-	-	-	62.2	37.8
19	Social and political relations	72.8	-	-	3.1	0.4	-	0.4	-	-	-	76.8	23.2
20	Warfare and hunting	67.1	10.9	2.3	0.4	0.4	-	0.4	-	-	-	81.4	18.6
21	Law	52.6	4.4	-	-	-	-	-	4.4	-	-	61.3	38.7
22	Religion and belief	36.5	9.6	5.8	-	-	-	7.7	-	-	3.8	63.5	36.5
23	Modern world	76.7	14.9	0.8	-	-	-	-	-	-	-	92.3	7.7
24	Miscellaneous function words	-	-	8.5	-	4.3	-	-	-	-	-	12.8	87.2
	all words	**50.8**	**4.3**	**2.4**	**1.7**	**1.7**	**0.8**	**0.6**	**0.1**	**0.1**	**0.2**	**62.7**	**37.3**

Table 5, analogous to Table 3 in §4.3, displays the proportions of selected diachronic layers of loanwords to all loanwords by semantic field.

Table 5: *Loanwords in Selice Romani by semantic field and diachronic layer (percentages)*

	Semantic field (field number)	Loans	Hungarian		Current L2s		Last 1000 years	
7	The house	92.7	79.3	−4.9	90.6	+1.9	100.0	+5.0
23	The modern world	92.3	95.9	+11.7	99.2	+10.5	100.0	+5.0
8	Agriculture and vegetation	89.8	82.1	−2.1	89.8	+1.1	98.1	+3.1
6	Clothing and grooming	86.7	82.2	−2.0	86.0	−2.7	88.6	−6.4
20	Warfare and hunting	81.4	84.3	+0.1	95.7	+7.0	100.0	+5.0
3	Animals	77.8	84.1	−0.1	95.1	+6.4	97.9	+2.9
19	Social and political relations	76.8	94.9	+10.7	94.9	+6.2	100.0	+5.0
1	The physical world	74.3	92.4	+8.2	92.4	+3.7	95.7	+0.7
22	Religion and belief	63.5	60.6	−23.6	72.8	−15.9	81.9	−13.1
18	Speech and language	62.2	92.0	+7.8	92.0	+3.3	100.0	+5.0
21	Law	61.3	85.8	+1.6	100.0	+11.3	100.0	+5.0
9	Basic actions and technology	60.5	87.9	+3.7	92.4	+3.7	97.9	+2.9
5	Food and drink	58.8	78.7	−5.5	80.9	−7.8	87.2	−7.8
14	Time	59.2	89.7	+5.5	89.7	+1.0	100.0	+5.0
4	The body	57.2	80.4	−3.8	82.6	−6.1	88.6	−6.4
10	Motion	57.6	86.5	+2.3	86.5	−2.2	95.4	+0.4
15	Sense perception	55.4	83.9	−0.3	83.9	−4.8	91.9	−3.1
16	Emotions and values	51.7	77.9	−6.3	77.9	−10.8	93.4	−1.6
11	Possession	51.2	77.5	−6.7	90.2	+1.5	94.5	−0.5
17	Cognition	50.7	100.0	+15.8	100.0	+11.3	100.0	+5.0
12	Spatial relations	47.7	82.4	−1.8	87.8	−0.9	93.3	−1.7
13	Quantity	37.0	51.9	−32.3	51.9	−36.8	93.2	−1.8
2	Kinship	32.1	80.7	−3.5	84.4	−4.3	91.9	−3.1
24	Misc. function words	12.8	0.0	−84.2	0.0	−88.7	49.4	−45.6
	Total	*62.7*	*84.2*	*0.0*	*88.7*	*0.0*	*95.0*	*0.0*

Disregarding the field *Miscellaneous function words* for the moment, we may observe the following: All fields contain from just below a third to over 90% loanwords. The overwhelming majority of fields contain more loanwords than non-loanwords (with the exception of *Kinship, Quantity*, and *Spatial relations*), and around a third of fields contain more than three quarters of loanwords. The

proportion of Hungarian loanwords to all loanwords ranges between a half and all in different semantic fields, with the bulk of fields showing more than three quarters of Hungarian loanwords. The proportions of loanwords from all current contact languages do not present a significantly different picture. At least four fifths of loanwords in any semantic field, and often all of them, have been borrowed within the last millenium. The fields that contain fewer loanwords in general also tend to contain, with some exceptions, a smaller proportion of the more recent, Greek and post-Greek, loanwords to all loanwords (since, however, the statistical significance of the proportions of different loanword layers will differ greatly for different fields, this latter observation should not be given too much weight).

There is certainly no single principle behind the ordering of the LWT semantic fields with regard to the proportion of loanwords they contain. Nevertheless, it may be observed that several fields consisting, to a considerable extent, of abstract concepts (e.g. *Quantity*, *Spatial relations*, *Cognition*, *Possession*, or *Emotions and values*) possess relatively low proportions of loanwords, whereas numerous fields that mostly contain very concrete meanings (e.g. *The house*, *Modern world*, *Agriculture and vegetation*, *Clothing and grooming,* or *Animals*) possess relatively high proportions of loanwords. Some of those semantic fields that stand out in Table 5 in various respects are discussed below:

- The field *The house* shows the highest proportion of loanwords. There are only three LWT meanings that must be expressed by an indigenous word: 'house', 'door', and 'to live, dwell' (< 'to sit').[18] It is likely that some loanwords in this field have been cultural insertions accompanying the speakers' sedentarization and other changes in their dwelling patterns and conditions (e.g. 'room'), although other loanwords have demonstrably replaced indigenous words (e.g. 'board') or pre-sedentarization loanwords

---

[18] In addition, there is an indigenous noun meaning 'space under one's head in bed' (whereas 'pillow' is a loanword), and two polysemous indigenous nouns that can be used to refer to 'floor' (primarily 'earth; land') and 'bed' (primarily 'place'), for both of which there are borrowed synonyms in the relevant specific meanings.

(e.g. 'stove'). It thus remains unclear to what extent extralinguistic factors can be made responsible for the extremely high proportion of loanwords in this semantic domain. The fact that this field consists almost exclusively of nouns, which are the most borrowable word class in Selice Romani (see §4.3), may also be significant.

- The second highest proportion of loanwords in the field *Modern world* is not surprising. Unlike The House, this field contains, expectedly, an above-average proportion of Hungarian and current loanwords. In fact, the only pre-Hungarian loanword in this field, *caklo* 'glass [material]; bottle' from South Slavic, has acquired its latter, modern-world, meaning through calquing the polysemy of the Hungarian noun *üveg*. In addition, there are a few relatively recent internal derivations in this field, and an indigenous noun meaning 'song', which is an ancient rather than modern concept in Romani culture.

- The field *Religion and Belief* stands out in showing the highest proportion of old, pre-Greek, loanwords. However, given that there are only three of them, viz. 'priest', 'witch' and 'sorcerer, wizard' (the field contains relatively few words in Selice Romani), their outstanding proportion is probably not statistically significant.

- The field *Quantity* has a relatively low proportion of loanwords and, especially, the lowest proportion of loanwords from Hungarian and from Selice Romani's current contact languages in general. Almost a half of quantity loanwords were borrowed from Selice Romani's previous European contact languages, viz. Greek and South Slavic, which otherwise contribute much smaller proportions of lexicon.

- The lowest proportion of loanwords is found in the field *Kinship*, although they still amount to almost a third of all Kinship words. Moreover, numerous expressions in this field are collocations containing a loanword or derivations from a loanword, and so the proportion of indigenous words is much lower. Indigenous kin terms that are used by all Selice Romani speakers are restricted to 'brother', 'sister', 'father', and 'mother' (the latter, however, may be a loanword). Only the older generations of Selice Romani speakers also use indigenous words for 'father-in-law', 'mother-in-law', and 'daughter-in-

law'. Further indigenous words in this field include 'human being', 'man; male' (which may be an old loanword), 'woman; female', and 'wedding'.

The semantic field that has by far the lowest proportion of loanwords, and which has been disregarded in the above discussion, are the *Miscellaneous function words*. There are only two loanwords here, one from Iranian ('without etc.') and one from Serbo-Croatian ('nothing'), i.e. none from Hungarian or any other current contact language. As a result, the various loanword proportions in this field are very different from those in all other fields. Note that this field only contains certain kinds of function words, including some of the less borrowable ones (e.g. demonstratives, basic adpositions, auxiliary verbs), and should not be considered representative of function words in general: the semantic word class of function words has more than three times as high proportion of loanwords (see §4.3).

5. Integration of loanwords

5.1. Phonological integration of loanwords

The phoneme inventory of Selice Romani is almost identical to that of the local dialect of Hungarian, partly because Selice Romani has both acquired and lost a number of phonemic distinctions due to contact with this contact language (cf. Elšík 2007+). The only Hungarian phonemes to get phonologically adapted in Selice Romani loanwords are the front rounded vowels: the mid /ö/ [ø] and /ő/ [ø:] and the high /ü/ [y] and /ű/ [y:]. They are mostly replaced with their front unrounded counterparts, the mid /e/ [e ~ æ] and /é/ [æ:] and the high /i/ [i] and /í/ [i:], respectively, e.g. Hungarian *csütörtök* 'Thursday' > Selice Romani *čitertek-o* and Hungarian *kőműves* 'bricklayer' > Selice Romani *kémíveš-i*. One systematic exception occurs in Selice Romani loanwords of polysyllabic Hungarian nouns whose base form ends in the long front rounded vowels. Here, Hungarian /ő/ and /ű/ are replaced with the back rounded vowels /ó/ [o:] and /ú/ [u:], respectively, e.g. Hungarian *kereskedő* 'merchant' > Selice Romani

kereškedó and Hungarian *kesztyű* 'glove' > Selice Romani *kestű(-va)*. However, when these nouns are parts of compounds in Hungarian, the regular unrounding applies, e.g. Hungarian *tüdő* > Selice Romani *tidó* 'lung' but Hungarian *tüdő*+*baj* [lung + trouble] > Selice Romani *tidébaj-a* 'pulmonary tuberculosis'. Also regular is the phonological adaptation in loanwords of Hungarian adjectivals and monosyllabic nouns ending in the long front rounded vowels, e.g. Hungarian *első* 'first' > Selice Romani *éšé-n-o*, Hungarian *könnyű* 'light; easy' > Selice Romani *keňňí-n-o*, Hungarian *fő* 'head; chief' > Selice Romani *fé* 'chief'. Note that there is no absolute constraint on word-final /é/ or /í/ in Selice Romani.[19]

Similarly, Slovak and Czech phonemes that are absent from Selice Romani (and Hungarian) must or may get phonologically adapted, e.g. optional [x > kʰ] in Slovak *východ* 'east' > Selice Romani *víkhod-o*, and obligatory [r̝ > ʃ] in Czech *pepř* > Selice Romani *pepš-o* 'black pepper'. Many apparent instances of phonological adaptation in current Selice Romani loanwords in fact reflect dialectal source forms, e.g. Selice Romani *čekíl-n-o* 'shallow' < Hungarian dialectal *csekíl*, cf. standard *sekély*; or adoption of the source language's non-base stem variants, e.g. Selice Romani *samar-a* 'donkey' < Hungarian *szamar-*, cf. the base stem *szamár*. In addition to these factors, post-contact phonological changes must also be taken into account when one tries to identify adaptation processes in older loanwords. For example, the Serbo-Croatian word *volja* 'will; mood' was probably borrowed without any phonological adaptation before it has changed to present-day Selice Romani *vója* 'good mood', due to regular Hungarian-induced phonological developments. One of the few clear instances of pre-Hungarian phonological adaptation is the change [y > u] in Early Romani *kurko*

---

[19] The regular unrounding of the front rounded vowels is also a characteristic ethnolectal feature of some Selice Romani speakers' Hungarian. Some Selice Romani–Hungarian bilinguals thus lack the front rounded vowels in both of their primary languages, while for others unrounding is an L1-internal adaptation process. In addition, there is some interesting lexical and sociolinguistic variation with regard to unrounding in the latter group of speakers: certain loanwords tend to retain the front rounded vowels, and some speakers tend to retain them in more loanwords than others. It seems that the lack of phonological adaptation in Selice Romani functions as a sociolinguistic marker of a kind of prestige associated with success in the non-Romani society.

'Sunday; week', a loanword of Medieval Greek *kyrikó(n)* 'Lord's (day); Sunday' (Tzitzilis 2001: 327).

5.2. Morphological integration of loanwords

Loanwords that are assigned the status of an inflected Selice Romani word class (noun, verb, or adjective) are, as a rule, morphologically integrated into Selice Romani inflectional patterns. However, there is a general division in Romani between two major diachronic layers of loanwords with regard to their degree of integration: loanwords from pre-Greek contact languages are fully integrated and indistinguishable from indigenous words on morphological grounds, whereas loanwords from post-Greek contact languages are, or can be reconstructed to have been in Early Romani, overtly marked by various morphological means as loanwords. Loanwords from Greek, which is the source of most loanword markers (e.g. Bakker 1997), are split between these two layers: some Hellenisms, presumably the early ones, are fully integrated, while others, presumably the later ones, are overtly marked as loanwords. This diachronic division is synchronically reflected as a morphologically encoded etymological compartmentalization of the lexicon: older loanwords, together with indigenous words, have what I term *oikoclitic* morphology, while more recent loanwords have *xenoclitic* morphology. The distinction between oikoclisis and xenoclisis, which can be reconstructed for Early Romani, has undergone a variety of analogical developments in individual Romani dialects, affecting not only individual lexemes, but also whole inflectional and derivational classes (see Elšík & Matras 2006: 324–333 for an overview).

The distinction between the full integration (oikoclisis) of earlier loanwords and marked integration (xenoclisis) of later loanwords is well retained in Selice Romani **noun** inflection. Xenoclitic loanwords are characterized by borrowed nominative suffixes, mostly of Greek origin, and by analogically reshaped oblique stem suffixes (see Elšík 2000, Matras 2002: 80–85 for details). For example, oikoclitic masculine loan-nouns in *-o* (e.g. *čár-o* 'bowl, dish' from Dravidian, *t̆rm-o* 'worm' from Persian, and *kurk-o* 'Sunday; week' from Greek)

take the indigenous nominative plural suffix *-e* and the indigenous oblique singular suffix *-es-*, whereas xenoclitic masculine loan-nouns in *-o* < Early Romani **-os* (e.g. *fór-o* 'town' from Greek, *prah-o* 'dust, powder' from South Slavic, *világ-o* 'world' from Hungarian, and *pepš-o* 'black pepper' from Czech) take the borrowed nominative plural suffix *-i* and the reshaped oblique singular suffix *-os-*. Other inflectional classes show different markers, but the principle remains the same.

Similarly, pre-Greek and early Greek loan-**verbs** show full morphological integration and are structurally indistinguishable from indigenous verbs. Post-Greek loan-verbs, on the other hand, are marked out by an overt (and dedicated) adaptation marker, the Greek-origin suffix *-in-*, which is added to an inflectional stem of the source verb (e.g. *vič-in-* 'to shout' from Serbo-Croatian *vič-*, *dógoz-in-* 'to work' from Hungarian *dolgoz-*), and followed by regular indigenous inflections. The suffix, which is a pre-inflectional though non-derivational morpheme, was extracted from lexical borrowings of Greek verbs with the present stem in *-in-*. Though none of these have been retained in Selice Romani, the suffix has been extended to those Greek loan-verbs that originally contained a different suffix, e.g. *rum-in-* 'to destroy, break, damage, spoil' from Greek *rim-az-* 'to ravage'. Dialect comparison suggests that the suffix *-in-* was originally specialized for non-perfective adaptation of some transitive loan-verbs in Romani (Matras 2002: 130). In Selice Romani, however, it has developed into a general, aspect- and valency-neutral, verb-adaptation marker.[20] Nonce loan-verbs from Slovak or Czech show a distinct pattern of morphological adaptation: their infinitive stems get adapted by the Hungarian-origin adaptation suffix *-ál-*,[21] in addition to the regular adaptation suffix *-in-*, e.g. *sledov-ál-in-* 'to observe, follow' from Slovak/Czech *sled-ov-a-*.

---

[20] The Greek-origin suffix **-(V)s-*, which appears to have been the marker of perfective adaptation of all loan-verbs and of non-perfective adaptation of intansitive loan-verbs (Matras 2002: 130), has acquired novel functions in Selice Romani (cf. Elšík 2007 +).

[21] Although Kenesei et al. (1998: 357–358) describe the Hungarian suffix *-ál-* as a de-nominal verb-deriving marker, their examples show that it is in fact a verb-adapting suffix, which is synchronically distinct from the de-nominal verb-deriving suffix *-(V)l*.

In **adjectives**, the distinction between xenoclitic and oikoclitic inflection, which is attested in most Romani dialects and reconstructable for Early Romani (e.g. Boretzky & Igla 2004: 112–113), has been lost due to internal analogical developments in all South Central dialects of Romani, including Selice Romani (cf. Elšík et al. 1999: 334, Elšík & Matras 2006: 329). All borrowed adjectives – i.e. not only those borrowed from Selice Romani's pre-Greek contact languages – now inflect exactly like indigenous adjectives and employ the former oikoclitic inflectional suffixes. In loanwords from pre-Hungarian contact languages, these inflections are suffixed directly to the inflectional stem of their source adjective, e.g. Selice Romani *žut-o* 'yellow' from Serbo-Croatian *žut*. In loanwords from Hungarian, on the other hand, the suffixation of the indigenous inflections to the source adjective's inflectional stem is mediated by overt and dedicated adaptation suffixes of South Slavic origin, e.g. Selice Romani *kík-n-o* 'blue', *keňňí-n-o* 'light; easy', or *sirk-av-o* 'grey' from Hungarian *kék*, *könnyű*, and *szürke*.[22] Like the verb-adapting suffix, both Selice Romani adjective-adapting suffixes, *-n-* and *-av-*, are pre-inflectional morphemes, since they are part of the inflectional stem of borrowed Selice Romani adjectives, though they are not derivational. While Selice Romani lost the original, Early Romani, marking of the original (pre-Greek vs. post-Greek) etymological compartmentalization in adjectives at some point of its history, it has developed a different kind of marking of a different (pre-Hungarian vs. Hungarian) etymological compartmentalization.

To sum up, there are three regular types of morphological integration of loanwords in Selice Romani: a) adaptation through unmarked (oikoclitic) inflectional integration with pre-Greek and early Greek nouns and verbs and with pre-Hungarian adjectives; b) adaptation through marked (xenoclitic) inflectional integration with late Greek and post-Greek nouns; c) (xenoclitic) adaptation by overt pre-inflectional suffixes with late Greek and post-Greek

---

[22] The distribution of the two adjective-adapting suffixes is conditioned by the weight of the source adjective's final syllable: Hungarian adjectives ending in a light syllable, i.e. in a short vowel, are adapted by the suffix *-av-*, which in addition triggers a deletion of the final vowel of the source form, whereas Hungarian adjectives ending in a heavy syllable, i.e. in a consonant (cluster) or a long vowel, are adapted by the suffix *-n-*.

verbs and with Hungarian adjectives; loan-verbs from Slovak and Czech stand out within this latter type in taking a morphologically complex adaptation marker. Only a few loanwords deviate from these regular patterns. To cite just one example: The Selice Romani noun *kóbás-kiň-a* 'a kind of sausage', from Hungarian *kolbász*, is adapted by means of the South Slavic suffix *-kiň-* plus the regular xenoclitic feminine inflection of Greek origin. This is quite curious since the former suffix is otherwise only used to derive feminine counterparts to masculine nouns denoting male humans (e.g. *šógor-kiň-a* 'sister-in-law' derived from *šógor-i* 'brother-in-law').

5.3. Speakers' attitudes to loanwords

Lexical variation between different generations of Selice Romani speakers shows that loanwords are entering the language at a relatively fast rate. There are several obsolete pre-Hungarian lexical expressions, which are familiar to, but not regularly used by, the oldest speakers and which have now been effectively replaced with loanwords from Hungarian, e.g. *kirivo* (< Kurdish) vs *náso* 'co-father-in-law', or *kárja d-* (indigenous) vs *levin-* 'to shoot'. There are also quite a few pre-Hungarian words, which are regularly used by older speakers but are usually replaced with Hungarian loanwords by younger speakers, e.g. *paraštíí* (< Greek) vs. *pinteko* 'Friday', or *tritóneste* (< Greek + indigenous inflection) vs *harmadikán* 'on the third (day of a month)'. My consultants never expressed any regret or compunction over the loss of the "old" words in the several discussions of lexical replacement I have provoked or witnessed, and the use of nonce loanwords from Hungarian (or Slovak or Czech) by Selice Romani speakers does not appear to be stigmatized in any way or viewed as "corruption" of the language. Selice Romani speakers often explain their group's self-designation as *ungrike Roma* 'Hungarian Roms' with reference to the presence of many "Hungarian words" in their variety of Romani.

However, the native concept of "Hungarian words," i.e. words that are recognized by the Selice Romani–Hungarian bilinguals as identical or similar in both of their primary languages, does not imply that their presence in Selice

Romani is automatically ascribed to borrowing from Hungarian. Several consultants have suggested to me that, alongside loanwords from Hungarian, there are also Romani words that are "simply similar" to Hungarian words without being loanwords.[23] This concept of ahistorical lexical similarity is likely to be connected to the native conceptualization of the group's history: the Hungarian Roms of Selice lack any narrative of external origin, claiming that they have lived in the village "from times immemorial."[24] Although the native criteria for distinguishing the two classes of Hungarianisms (and the extent to which this distinction is actually shared in the community) remain to be investigated, it seems that loanwords from Hungarian that are used across all generations and regularly employed in Romani discourse are not considered to be loanwords. Though they are referred to as "Hungarian words" in some contexts, in other contexts the speakers describe them as "proper Romani words." This appropriation strategy is likely to be linked to the speakers' tolerance for lexical borrowing.

There are few productive onomasiological processes within Selice Romani and the language relies heavily on loanwords in creating new naming units, especially in nouns. Unlike some Romani varieties that employ internal word-formation processes to create a layer of secret vocabulary in certain semantic domains (cf. Matras 2002: 223), Selice Romani does not seem to avoid loanwords in these domains. For example, while in most Romani varieties the regular word for 'policeman' is a Romani-internal formation that is not

---

[23] My consultants have never mentioned the third possibility, viz. that some lexical similarity between Selice Romani and Hungarian can be ascribed to borrowing from the former language into the latter. Nevertheless, a few words do show this kind of history, e.g. Selice Romani *péro* 'Romani settlement' > local Hungarian *péró* (> local Slovak *pérov*).

[24] The academic theory of an Indian origin of the Roms is known to some Roms from mass media and Romani(-related) publications, but it does not seem to enjoy any special status among the various hypotheses proposed by outsiders (such as that the Roms originate in Egypt, Palestine, Romania, or Spain).

comprehensible to outsiders, it is a loanword from Hungarian in Selice Romani: *čendéri < csendőr* 'gendarm'.[25]

6. Grammatical borrowing

Selice Romani has been affected by grammatical borrowing to a great extent.[26] Due to space limitations I will only present a very brief summary here (see Elšík 2008, for a more detailed overview). Several types of grammatical borrowing are distinguished below. First, Selice Romani has borrowed various kinds of contact language function words, only some of which are represented in the LWT-based subdatabase. Next, there are a number of borrowed affixes in Selice Romani. (I distinguish between affix *copying*, which is the direct transfer of contact language affixes without the mediation of lexical borrowing, and affix *extraction*, which consists in importation of contact language affixes within morphologically complex loanwords and their subsequent analogical extension to bases that do not originate in the source language of the affixes.) Finally, Selice Romani frequently replicates source language morphosyntactic patterns (constructions and categories) without necessarily borrowing the actual contact language morphemes that encode these patterns. The following summary only takes into account the post-Indian stages of the language (see especially Emeneau (1956) and Masica (1976) for grammatical convergence in South Asia and Friedman 2000: 95–6 for a brief overview of traces of South Asian areal features in Romani):

---

[25] To my knowledge only two related nouns, the masculine *čačuno* and feminine *čačuni* (derivations of the indigenous adjective *čáčo* 'true'), serve cryptolalic functions in Selice Romani: they may be used to refer to any human referent in situations when the referent and/or by-standers are not supposed to understand that the referent is being talked about.

[26] Selice Romani, together with Guarani, shows the greatest extent of grammatical borrowing among the 25 languages included in a recent cross-linguistic survey (Matras & Sakel 2008), exhibiting some kind of contact influence in 31 out of 36 prominent structural domains (Matras 2008).

1. **Function words** (selective): *Iranian*: possibility modal; *Greek*: several numerals and quantifiers; address particle; temporal deictic particle; *South Slavic*: several quantifiers; distributive numeral particle; optative/permissive particle; negative scalar focus particle and contrastive negative coordinator; impersonal negative pronoun; *Hungarian*: most coordinators; many adverbial subordinators; factual complementizer; question particle in embedded polar questions; generic obligation modal; several quantifiers and degree words; numerous preverbs; a few marginal postpositions; several adverbial deictics, interrogatives, indefinites and reflexives; repetition, utterance-level and phasal adverbs; focus particles; affirmative answer particles; interjections; fillers; sequential discourse markers; and more.

2. **Affix copying**: *South Slavic*: negative marker in negative pro-words; *Hungarian*: superlative marker; deictic-identity and deictic-contrast markers in demonstratives; specific indefinite, free choice indefinite and universal-quantification markers in pro-words.

3. **Affix extraction**: *Iranian*: comparative marker; *Greek*: nominative noun inflections; passive participle marker; loan-verb adaptation marker; markers deriving relational adjectives, ethnic adverbs, and ordinal numerals; *South Slavic*: loan-adjective adaptation markers; markers deriving feminine human nouns and attenuative adjectives; *Hungarian*: infinitive inflection; loan-verb adaptation marker; markers productively deriving action and artificial nouns, active de-verbal adjectives, de-nominal and causative verbs, and similative adverbs; numerous unproductive derivational markers.

4. **Morphosyntactic replication** (selective): *Middle Eastern*: development of interrogative-based relativizers; reduction of non-finite constructions; remoteness marking in verbs; *Greek*: development of a proclitic definite article; emergence of prepositions; shift to a basic predicate–object order; *South Slavic*: de-interrogative structure of negative pro-words; negative agreement with negative pro-words; *Hungarian*: morphemic structure of the reciprocal pronoun; syntactic category of preverbs; morphological categories of associative plurals in nouns and of frequentatives in verbs; elaboration of the morphological category of orientation in spatial adpositions and pro-words; reduction of gender in anaphoric pronouns and of feminine derivation

in nouns denoting animals; retention and productivity of the morphological categories of degree in adjectives and of causatives in verbs; subjunctive-based infinitive construction; encoding of various case relations; absence of case agreement in numeral constructions; negation of phasal adverbs; ontological restrictions on relativizers; certain pragmatic and syntactic aspects of linear constituent order; and more.

The above summary shows that Hungarian, the primary current contact language, is by far the most important source of all types of grammatical borrowings in Selice Romani. This is in line with the role of Hungarian as the most important source of loanwords in Selice Romani. In contrast, Slovak and Czech, the secondary current contact languages, appear not to have exerted any grammatical influence on Selice Romani. Of the past contact languages of Selice Romani, Greek stands out as a major source of grammatical borrowings, which contrasts with the relatively low proportion of lexical Hellenisms. Recall, however, that the Greek loanwords that have been retained in Selice Romani represent only a small fraction of all the Greek loanwords that are reconstructable for Early Romani (see §3.5).

7. Conclusions

Selice Romani, an Indo-Aryan language of Slovakia, has been in contact with, and has borrowed words from, a number of different languages in the course of its history. Loanwords amount to almost two thirds of those lexicalized meanings that are sampled by the Loanword Typology meaning list. Loanwords of all major word classes and all semantic fields are well represented in Selice Romani lexicon, although certain classes of function words are not borrowable and although it seems that relatively more abstract semantic fields are less affected by lexical borrowing. Nouns have a greater proportion of loanwords than any other word class. Selice Romani makes use of established and productive morphological mechanisms to integrate inflected loanwords, including nonce loanwords from current contact languages. Instances of grammatical borrowing

of both matter and pattern are abundant. Hungarian, a language in which speakers of Selice Romani have now been bilingual for many generations if not several centuries, is far and away the most important source of loanwords and grammatical borrowings. In contrast, Slovak and Czech, the other current contact languages of Selice Romani, contribute relatively few loanwords and no grammar.

The general sociolinguistic situation of all Romani varieties is highly favourable to contact-induced developments, since all adult Romani speakers are bilingual in the relatively prestigious languages of the dominant populations and since, at the same time, Romani linguistic ideologies are tolerant of linguistic borrowing, including borrowing of linguistic matter, which is less difficult to monitor and control. Nevertheless, there are differences between individual Romani varieties with regard to the degree of contact influences, which, in addition to length of contact with a particular contact language, reflect differences in sociolinguistic situations. The long-settled Roms of Hungary and the Hungarian regions of Slovakia have developed a strong orientation towards Hungarian cultural models. While in most communities of Hungarian Roms this cultural orientation has contributed to language shift from Romani to Hungarian, in the few extant varieties of Hungarian Rumungro, including Selice Romani, it has facilitated the propagation of Hungarian-induced linguistic innovations in Romani. The use of nonce loanwords does not appear to be stigmatized in any way and the departing "old" words, whose gradual replacement is observable across generations, are not mourned by the speakers. The acceptance and introduction of lexical Hungarianisms appears to be an overt expression of the amalgamated ethnic identity of the Hungarian Roms of Selice.

Acknowledgements

References

Ascoli, Graziadio I. 1865. *Zigeunerisches*. Halle: Heynemann.

Bakker, Peter. 1997. "Athematic morphology in Romani: The borrowing of a borrowing pattern." Matras, Yaron, Peter Bakker & Hristo Kyuchukov (eds.) *The typology and dialectology of Romani.* (= Current Issues in Linguistic Theory, 156.) Amsterdam: John Benjamins. 1–21.

Bakker, Peter. 1999. Review of Mānušs *et al.* 1997. *Grazer Linguistische Studien* 51: 167–170.

Beníšek, Michael. 2006. "Ke kořenům slova *rom*." [On the roots of the word *rom*.] *Romano džaniben*, jevend, 9–28.

Berger, Hermann. 1959. "Die Burušaski-Lehnwörter in der Zigeunersprache." *Indo-Iranian Journal* 3: 17–43.

Boretzky, Norbert. 1995. "Armenisches im Zigeunerischen (Romani und Lomavren)." *Indogermanische Forschungen* 100: 137–155.

Boretzky, Norbert. 1999. "Die Gliederung der Zentralen Dialekte und die Beziehungen zwischen Südlichen Zentralen Dialekten (Romungro) und Südbalkanischen Romani-Dialekten." In: Halwachs & Menz 1999, 210–276.

Boretzky, Norbert. 2003. *Die Vlach-Dialekte des Romani. Strukturen, Sprachgeschichte, Verwandtschaftsverhältnisse, Dialektkarten*. Wiesbaden: Harrassowitz.

Boretzky, Norbert. No date. *Slavismen im Romani und der Fall* beda/bedo. Manuscript.

Boretzky, Norbert & Igla, Birgit. 1991. *Morphologische Entlehnungen in den Romani-Dialekten*. (= Arbeitspapiere des Projektes "Prinzipien des Sprachwandels", 4.) Essen: Fachbereich Sprach- und Literaturwissenschaften an der Universität Essen.

Boretzky, Norbert & Igla, Birgit. 1994. *Wörterbuch Romani–Deutsch–Englisch für den südosteuropäischen Raum: mit einer Grammatik der Dialektvarianten*. Wiesbaden: Harrassowitz.

Boretzky, Norbert & Igla, Birgit. 2004. *Kommentierter Dialektatlas des Romani*. Teil 1: *Vergleich der Dialekte*. Teil 2: *Dialektkarten mit einer CD Rom*. Wiesbaden: Harrassowitz Verlag 2004.

Briggs, George Weston. 1953. *The Doms and their near relations*. Mysore: Wesley Press.

Buck, Carl Darling. 1949. *A dictionary of selected synonyms in the principal Indo-European languages*. Chicago: University of Chicago Press.

Burrow, Thomas. 1945. "Some Dravidian words in Sanskrit." *Transactions of the Philological Society*, 79–120.

Burrow, Thomas. 1946. "Loanwords in Sanskrit." *Transactions of the Philological Society*, 1–30.

Burrow, Thomas. 1947–1948. "Dravidian studies VII: Further Dravidian words in Sanskrit." *Bulletin of the School of Oriental and African Studies* 12, 365–396.

Burrow, Thomas. 1973 (3rd edition). *The Sanskrit language.* London: Faber and Faber.

Burrow, Thomas & Emeneau, Murray B. 1960, ²1984. *A Dravidian etymological dictionary*. Oxford: Clarendon Press.

Elšík, Viktor. 2000. "Romani nominal paradigms: Their structure, diversity, and development." In: Elšík, Viktor & Matras, Yaron (eds.) *Grammatical relations in Romani: The noun phrase*. Amsterdam: Benjamins. 9–30.

Elšík, Viktor. 2003. "Interdialect contact of Czech (and Slovak) Romani varieties." *International Journal of the Sociology of Language* 162: 41–62.

Elšík, Viktor. 2006. Review of Boretzky & Igla 2004. *Romani Studies* 5, 16: 105–111.

Elšík, Viktor. 2007+. "Grammatical borrowing in Hungarian Rumungro." In: Matras & Sakel, 2007+.

Elšík, Viktor, Hübschmannová, Milena & Šebková, Hana. 1999. "The Southern Central (*ahi*-imperfect) Romani dialects of Slovakia and northern Hungary." In: Halwachs & Menz 1999, 277–390.

Elšík, Viktor & Matras, Yaron. 2006. *Markedness and language change: the Romani sample*. Berlin: Mouton de Gruyter.

Emeneau, Murray B. 1956. "India as a linguistic area." *Language* 32: 3–16.

Fraser, Angus. 1992. *The Gypsies*. Oxford: Blackwell.

Friedman, Victor A. 1986. "Romani *te* in a Balkan context." *Językowe studia bałkanistyczne* 1: 39–48.

Friedman, Victor A. 1988. "A Caucasian loanword in Romani." In: DeSilva, Cara, Grumet, Joanne & Nemeth, David J. (eds.) *Papers from the eighth and ninth annual meetings, Gypsy Lore Society, North American Chapter*. New York: Gypsy Lore Society, North American Chapter. 18–20.

Friedman, Victor A. 1989. "Toward defining the position of Turkisms in Romani." In: Balić, Sait *et al.* (eds.) *Jezik i kultura Roma*. Sarajevo: Institut za proučavanje nacionalnih odnosa. 251–267.

Friedman, Victor A. 2000. "Romani in the Balkan Linguistic League." In: Tzitzilis, Christos & Symeonidés, Kh. (eds.) *Valkanikē glōssologia: sygxronia kai diaxronia. Balkanlinguistik: Synchronie und Diachronie.* Thessaloniki: University of Thessaloniki. 95–105.

Grant, Anthony P. 2003. "Where East meets West: observations on a list of Greek loans of European Romani." In: Grant, Anthony P. (ed.) *Papers in contact linguistics.* (Interface, Bradford Studies in Language, Culture and Society, 6). Bradford: Department of Languages and European Studies, University of Bradford. 27–69.

Halwachs, Dieter W. & Menz, Florian (eds.) 1999. *Die Sprache der Roma. Perspektiven der Romani-Forschung in Österreich im interdisziplinären und internationalen Kontext.* Klagenfurt: Drava.

Hancock, Ian. 1987. "Il contributo armeno alla lingua romani." *Lacio Drom* 23: 4–10.

Hancock, Ian. 1995. "On the migration and affiliation of the Ḍōmba: Iranian words in Rom, Lom and Dom Gypsy." In: Matras, Yaron (ed.) *Romani in contact: the history and sociology of a language.* Amsterdam: Benjamins. 25–51.

Kenesei, István, Robert M. Vago, and Anna Fenyvesi. 1998. *Hungarian.* London: Routledge.

Kostov, Kiril. 1996. Review of Boretzky & Igla 1994. *Grazer Linguistische Studien* 46: 131–142.

Kuiper, Franciscus B. J. 1948. *Proto-Munda words in Sanskrit.* Amsterdam: N. V. Noord-Hollandsche Uitgevers Maatschappij.

Kuiper, Franciscus B. J. 1991. *Aryans in the Rigveda.* Amsterdam: N. V. Noord-Hollandsche Uitgevers Maatschappij.

Lubotsky, Alexander M. 2001. "The Indo-Iranian substratum." In: Carpelan, Chr., Parpola, A. & Koskikallio, P. (eds.) *Early contacts between Uralic and Indo-European: linguistic and archaeological considerations. Papers presented at an international symposium held at the Tvärminne Research Station of the University of Helsinki 8-10 January 1999.* (Mémoires de la Société Finno-ougrienne 242.) Helsinki 2001. 301–317.

Mānuš, Léksa. 1994. "Le etimologie di venti vocaboli zingari." *Lacio Drom* 30: 26–34.

Mānušs, Leksa, Neilands, Jānis & Rudevičs, Kārlis. 1997. *Čigānu–latviešu–angļu etimoloģiskā vārdnīca un latviešu–čigānu vārdnīca.* [Gypsy–Latvian–English etymological dictionary and Latvian–Gypsy dictionary.] Rīgā: Zvaigzne ABC.

Masica, Colin P. 1976. *Defining a linguistic area: South Asia.* Chicago: University of Chicago Press.

Masica, Colin P. 1979. "Aryan and non-Aryan elements in North Indian agriculture." In: Deshpande, Madhav M. & Hook, P. E. (eds.) *Aryan and Non-Aryan in India.* Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan. 55–151.

Matras, Yaron. 1994. "Structural Balkanisms in Romani." In: Reiter, Norbert, Hinrichs, Uwe & van Leeuwen-Turnovcová, Jiřina (eds.) *Sprachlicher Standard und Substandard in Südosteuropa und Osteuropa.* Wiesbaden: Harrassowitz. 195–210.

Matras, Yaron. 1995. "Verb evidentials and their discourse function in Vlach Romani narratives." In: Matras, Yaron (ed.) *Romani in contact: The history, structure and sociology of a language.* Amsterdam: Benjamins. 95–123.

Matras, Yaron. 1996. Review of Boretzky & Igla 1994. *Zeitschrift für Balkanologie* 32: 214–224.

Matras, Yaron. 1999. "Johann Rüdiger and the study of Romani in eighteenth-century Germany." *Journal of the Gypsy Lore Society, Fifth Series* 9: 89–106.

Matras, Yaron. 2002. *Romani: a linguistic introduction.* Cambridge: Cambridge University Press.

Matras, Yaron. 2005. "The classification of Romani dialects: a geographic–historical perspective." In: Schrammel, Barbara, Halwachs, Dieter W. &

Ambrosch, Gerd (eds.) *General and applied Romani linguistics: Proceedings from the 6th International Conference on Romani Linguistics*. Munich: Lincom Europa. 7–22.

Matras, Yaron. 2007+. "The borrowability of structural categories." In: Matras & Sakel, 2007+.

Matras, Yaron & Sakel, Jeanette (eds.) 2007+. *Grammatical borrowing in cross-linguistic perspective*. Berlin: Mouton de Gruyter. In press.

Mayrhofer, Manfred. 1986–2001. *Etymologisches Wörterbuch des Altindoarischen*. 3 volumes. Heidelberg: Carl Winter.

Messing, Gordon M. 1988. *A glossary of Greek Romany: as spoken in Agia Varvara (Athens)*. Columbus, Ohio: Slavica.

Mijatović, Boško. 2006. *Tobacco and the Serbian state in the 19th century*. Center for Liberal-Democratic Studies 2006. http://www.clds.org.yu/newsite/Duvan-e.pdf

Miklosich, Franz. 1872–1880. *Über die Mundarten und Wanderungen der Zigeuner Europas*. Volumes I-XII. Wien: Karl Gerold's son.

Pobożniak, Tadeusz. 1964. *Grammar of the Lovari dialect*. (= Prace komisji orientalistycznej, 3.) Kraków: Państwowe wydawnictwo naukowe.

Pott, August F. 1844–1845 [1964]. *Die Zigeuner in Europa und Asien. Ethnographisch-linguistische Untersuchung vornehmlich ihrer Herkunft und Sprache*. Halle: Heynemann. [Leipzig: Edition Leipzig].

Sampson, John. 1926 [1968] *The dialect of the Gypsies of Wales, being the older form of British Romani preserved in the speech of the clan of Abram Wood*. Oxford: Clarendon Press.

Soravia, Giulio. 1988. "Di alcune etimologie zingariche." *Archivio glottologico italiano*, 73: 3–11.

Soulis, George. 1961. "The Gypsies in the Byzantine Empire and the Balkans in the Late Middle Ages." *Dumbarton Oaks Papers* 15: 142–165.

Southworth, Franklin C. 2005a. *Linguistic archaeology of South Asia*. London: Routledge-Curzon.

Southworth, Franklin C. 2005b. "Prehistoric implications of the Dravidian element in the NIA lexicon, with special reference to Marathi." *International Journal of Dravidian Linguistics* 34(1): 17–28.

Tálos, Endre. 1999. "Etymologica Zingarica." *Acta Linguistica Hungarica* 46: 215–268.

Toropov, Vadim G. 2004. *Istorija i fol'klor krymskix cygan.* [The history and folklore of Crimean Gypsies.] Moskva: Rossijskij naučno-issledovatel'skij institut kul'turnogo i prirodnogo nasledija im. D. S. Lixačeva.

Turner, Ralph L. 1924. *traš-* 'to frighten'. *Journal of the Gypsy Lore Society, Third series* 3: 38–43.

Turner, Ralph L. 1926. "The position of Romani in Indo-Aryan." *Journal of the Gypsy Lore Society, Third series* 5: 145–189.

Turner, Ralph L. 1962–1966. *A comparative dictionary of the Indo-Aryan languages.* Oxford: Oxford University Press.

Tzitzilis, Christos. 2001. "Mittelgriechische Lehnwörter im Romanes." In: Igla, Birgit & Stolz, Thomas (eds.) *"Was ich noch sagen wollte..." A Multilingual Festschrift for Norbert Boretzky on the Occasion of His 65th Birthday (Sprachtypologie und Universalienforschung, Supplements, Studia typologica 2).* Berlin: Akademie Verlag. 328–340.

Valtonen, Pertti. 1972. *Suomen mustalaiskielen etymologinen sanakirja.* [An etymological dictionary of the Finnish Gypsy language.] Helsinki: Suomalaisen kirjallisuuden seura.

Vekerdi, József. 1981. "On the social prehistory of the Gypsies." *Acta Orientalia Academiae Scientiarum Hungaricae* 23: 243–254.

Vekerdi, József. 1983. *A magyarországi cigány nyelvjárások szótára.* [A dictionary of Gypsy dialects of Hungary.] Pécs: Janus Pannonius Tudományegyetem Tanárképző Kara.

Vekerdi, József; with the assistance of Zsuzsa Várnai. 2000. *A comparative dictionary of Gypsy dialects in Hungary. Gypsy–English–Hungarian dictionary with English to Gypsy and Hungarian to Gypsy word lists.* Budapest: Terebess Publications. [revised edition of Vekerdi 1983]

Witzel, Michael. 1999a. "Substrate languages in Old Indo-Aryan (Rgvedic, Middle and Late Vedic)." *Electronic Journal of Vedic Studies* 5: 1–67.

Witzel, Michael. 1999b. "Early sources for South Asian substrate languages." *Mother Tongue,* Special Issue, October 1999, 1–70.

Witzel, Michael. 1999c. "Aryan and Non-Aryan names in Vedic India. Data for the linguistic situation, c. 1900–500 B.C." In: Bronkhorst, Johannes & Deshpande, Madhav M. (eds.) *Aryan and Non-Aryan in South Asia: Evidence, interpretation and ideology. Proceedings of the Michigan-Lausanne International Seminar on Aryan and Non-Aryan in South Asia, University of Michigan, Ann Arbor, 25-27 October 1996*. (= Harvard Oriental Series, Opera Minora, 3.) Cambridge: Department of Sanskrit and Indian Studies, Harvard University. 337–404.

Witzel, Michael. 2003. "Linguistic evidence for cultural exchange in prehistoric western Central Asia". *Sino-Platonic Papers* 129: 1–70.

Wolf, Siegmund A. 1960 [1987, 1993]. *Großes Wörterbuch der Zigeunersprache*. Mannheim [Hamburg: Buske].

Woolner, Alfred C. 1915. "Studies in Romani philology I: Personal pronouns". *Journal of the Gypsy Lore Society, Second series*, 9: 119–128.

Woolner, Alfred C. 1928. "Asoka and the Gypsies." *Journal of the Gypsy Lore Society, Third Series* 7: 108–111.

Appendix

Non-Indo-European Central Asian

búko	'liver'
sú	'needle'
rašaj	'priest'

Non-Indo-Aryan Indian

* (Para/Proto-)Munda; † Dravidian

čik *†	'mud'
čhá *	'Romani young man or boy; son; child'
rom *	'Rom; married Romani man; husband'
pro *	'foot, leg'
bango *	'crooked, bent, curved; lame'
čáro †	'bowl, dish'
čiken *†	'grease, fat'
tȋral †	'curd, quark, cottage cheese'
urďen †	'to put on, dress'
harno *	'short'
párno *	'white'
kálo †	'black'

Persian

* may also be Kurdish

véš	'woods, forest'
tȋrmo	'worm'
zár	'body hair, pubic hair; animal hair'
dumo	'back'
angušť	'finger, toe'
pór	'feather'
rezdan	'to shiver, tremble'
koro *	'blind'
kúči	'cup, mug'

mol	'wine'
angrusti	'ring'
resen	'to arrive; reach; fit into'
kamen	'to wish; want; love, like; owe'
bast	'luck, good luck'
zijand	'damage'
bi	'without; instead of; except for; because of'
patavo	'foot-rag, foot cloth'
hurdo	'tiny, petite, minute, small'

Kurdish

tover	'axe'

Ossetic

* may also be Persian

holev	'trousers'
verda *	'cart, wagon, carriage; pram, buggy'

Armenian

* may also be Persian

gra	'horse'
čekat *	'forehead'
burňik	'palm of the hand; handful'
t̃rhaj	'boot, high boot'
dudum	'pumpkin, squash, gourd'
kotor	'piece'
čoháni	'witch'
čoháno	'sorcerer, wizard'
humer *	'boiled or baked dough; pastry; noodle(s)'

Greek

papu	'grandfather'
papiň	'goose'

cipa	'skin; hide; leather; bark; coat, peel, shell'
kokal	'bone'
cimbla	'eyebrow'
zumi	'soup'
sapuňi	'soap'
kafidi	'table'
kopana	'trough'
irinen	'to turn'
drom	'way; road; path; journey'
ruminen	'to destroy, break; damage; spoil'
efta	'seven'
ofto	'eight'
eňňa	'nine'
trito	'third'
táha	'tomorrow'
kurko	'Sunday; week'
paraštú	'Friday'
hóli	'anger'
troman	'to dare, venture'
fóro	'town'
kopaj	'walking stick; club'
luluďi	'flower'
amoňi	'anvil'

South Slavic

* must be Serbo-Croatian ; † cannot be Serbo-Croatian

praho	'dust; powder'
nebo	'sky; heaven'
baba	'grandmother'
gerkáňi	'larynx, throat'
sléžinka	'spleen'
péťa *	'oven'
žila	'vein, artery; sinew, tendon'

mlíno	'mill'
klúčo	'key'
vodro †	'bed'
perníca	'pillow'
járko	'ditch'
bobo	'maize/corn'
duhano *	'tobacco'
klinco	'nail'
caklo *	'glass; bottle'
plasta	'canvas, awning, sheet; sail'
uze *	'beside, by, next to'
dugo *	'long'
cilo *	'whole; all'
dosta	'enough'
sobota †	'Saturday'
zeleno	'green'
žuto *	'yellow'
gizdavo	'proud, haughty'
erďavo *	'bad; evil; wrong'
vičinen *	'to shout, cry out'
priminen	'to promise'
molinen	'to pray'
ništa *	'nothing'
tresinen	'to shake'
smírom *	'in peace'

Hungarian

* may also be Slovak and/or Czech; † may also be South Slavic

világo	'world'
heďo	'mountain'
dombo	'hill'
igeňeššígo	'directness; rectitude; plain'
sigeto	'island'

parto	'coast, shore'
barlango	'cave'
tengeri	'sea'
ňugotno	'calm'
zavarošno	'rough (about sea); obscure; confused, mad, insane'
habo	'foam'
lagúna *	'lagoon'
hullámo	'wave'
fojó	'river'
patako	'brook, creek, stream'
forgó (1)	'whirpool'
močára	'swamp'
ešíši	'waterfall'
kevečo	'small stone; gravel'
feldrengíši	'earthquake'
hódačka	'moon, little moon'
čillaga	'star'
villámo	'(flash of) lightning; bolt of lightning, thunderbolt'
villámláši	'lightning'
zengíši	'thunder'
zivatalo	'storm'
sivárváňi	'rainbow'
világšágo	'light'
világoššágo	'light'
šetítšígo	'darkness'
árňíko	'shade, shadow'
dero	'hoarfrost; dew'
levegó(va)	'air'
felhó(va)	'cloud'
kedo	'fog'
jego	'ice'
faďinen	'to freeze'
lánga	'flame'

fišto	'smoke'
gézo	'steam'
hamu	'ash'
izgó(va)	'embers'
ďújtó	'match, lighter'
tata	'father!' [address form]
mama	'mother!' [address form]
sileji	'parents'
teštvírno	'sibling'
kettéšno	'twin'
onoka	'grandson; grandchild' [masculine]
onoka	'granddaughter' [feminine]
onokateštvírno	'cousin (male or generic)'
onokateštvíro	'male cousin'
apóši	'father-in-law'
aňóša	'mother-in-law'
aňóškiňa	'mother-in-law'
vejo	'son-in-law'
meňečke	'daughter-in-law'
šógori	'brother-in-law; sibling-in-law'
árvasto	'orphan'
özvedni	'widow'
özvedno	'widower'
rokoňi	'relative, kin'
rokoňšágo	'relatives, kinship'
čaládo	'family; child, immediate descendant'
áloto	'animal; livestock'
legeló	'pasture'
ištálló(va)	'stable, stall'
bika	'bull; ox'
borjúko	'calf'
báránka	'lamb, small sheep'
kano	'male pig; wild boar'

kečke	'goat (generic)'
čéderi	'stallion'
čikó	'foal, colt'
samara	'donkey'
kokaši	'cock, rooster'
čibóka	'chicken'
káča	'duck'
ficko	'nest'
bérmadara	'bat'
papagáji *	'parrot'
varjú(ka)	'crow'
galamba	'pigeon, dove'
čuviko	'owl'
mačka (1) *†	'cat'
pociko	'mouse'
patkáňi	'rat, sewer-rat'
haja	'scale'
čiga	'(cockle-)shell'
delfíno *	'dolphin'
farkaši	'wolf'
orosláňa	'lion'
medve	'bear'
róka	'fox'
sarvaši	'deer'
majmo	'monkey'
elefanto	'elephant'
púpošteve	'camel'
bugari	'beetle; insect'
švábo *	'cockroach'
hanďa	'ant'
póko	'spider'
pókháló	'spider web'
míhečke	'bee'

fullánka	'bee, wasp or a similar stingy insect'
daráža	'wasp'
súňoga	'mosquito'
kullánča	'tick'
kijó	'snake'
mókuši	'squirrel'
jaguári *	'jaguar'
lepke	'butterfly'
čigabiga	'snail'
bíka	'frog'
ďíko	'lizard'
tekňéšbíka	'turtle'
tešto	'body'
hajsálo	'hair'
sakálla	'beard'
korpa	'dandruff'
era	'vein, artery'
ódalbordo	'rib'
sarva	'horn'
farka	'tail'
hátgerinci	'spine'
halántíko	'temples'
kopoňa	'skull'
állarckapča	'jaw'
álla	'chin'
semeldeko	'eyebrow; eyelid'
sempilla	'eyelash'
pillogatinen	'to blink, twinkle'
kačingatinen	'to wink, give a sign by blinking; flirt'
turňa	'nostril'
íňa	'gums'
ňak(a)čiga	'nape of the neck; neck vertebra'
vállo	'shoulder'

lapocka	'shoulderblade'
keňeka	'elbow'
čukló	'wrist'
combo	'thigh; haunch'
térďa	'knee'
boka (1)	'ankle'
šarka	'corner; edge; heel (body part)'
ňoma	'footprint'
lípíši	'step, footstep; footprint'
sárňa	'wing'
mello	'chest, bosom; bust'
puppa	'navel'
tidó	'lung'
májo	'liver'
vešó	'kidney'
ďomra	'stomach'
derko	'waist'
forgó (2)	'hip (bone)'
míha	'womb'
lílegzinen	'to breathe'
ášítozinen	'to yawn'
čuklinen	'to hiccough'
keheginen	'to cough'
triskinen	'to sneeze'
iddzadinen	'to perspire'
ňálozinen	'to wet with saliva, lick'
horšoginen	'to snore'
firdinen	'to bathe'
íleto	'life'
ďilkolinen	'to murder'
dego	'carcass'
temetinen	'to bury'
šíra	'grave'

ďengavo	'weak'
lázo	'fever'
nátha *	'cold'
šebo	'wound, sore'
belešeto	'accident; injury'
kaparinen	'to scratch'
hólaga	'blister'
himló	'small-pox, pox, pock, boil'
geňňo	'pus'
jelo	'bodily mark', esp. 'scar'
orvoši	'physician; doctor (degree)'
orvoššágo	'medicine'
ďócceri	'medicine'
mirgo	'poison'
fárotno	'tired'
pihelinen	'to rest'
luštavo	'lazy'
kopasno	'bald'
bénavo	'lame'
nímavo	'mute'
íretno	'ripe'
pošvatno	'rotten'
ráginen	'to chew, champ, gnaw, munch, nibble'
žuvačkázinen	'to chew chewing gum'
kemence	'oven'
šerpeňó(va)	'stew-pan, shallow pot'
kávéfézó	'kettle for making coffee'
palacinkašító	'pan'
táňíri	'plate'
koršó	'jug, pitcher'
vella	'fork; pitchfork'
fogó	'tongs'
reggeli	'breakfast'

ebído	'lunch'
vačora	'supper'
sitálinen	'to sieve, sift'
pucolinen	'to scrape'
keverinen	'to mix, stir'
kovási	'dough'
dagastinen	'to knead'
ďúrinen	'to knead or roll (dough); crush'
élinen	'to grind'
kóbáskiňa	'(a kind of) sausage'
zéčígo	'vegetables'
babo	'bean'
krumpja	'potato'
ďiméčo	'fruit'
čomó	'knot; knob; bundle, bunch, cluster; batch'
figa	'fig'
sélló(va)	'grapel; vine'
oliva *	'olive'
olaji	'oil'
paprika *	'paprika, pepper'
míza	'honey'
cukro *	'sugar'
šajto	'cheese'
italo	'drink', esp. 'alcoholic drink'
šero	'beer'
tojáši	'egg'
sabó	'tailor'
varó	'tailor'
aňago	'material, matter, stuff; cloth'
ďapjúšno aňago	'wool'
ďapjúaňago	'wool'
vásoňi	'linen; canvas'
pamuko †	'cotton'

šejmo	'silk'
bunda	'fur; fur coat'
ketinen	'to bind; plait, brad; knit; weave'
cérna	'thread'
feštinen	'to dye; paint; put on make up'
palášťa	'cloak'
keppeňi	'raincoat, macintosh'
kabáto *	'coat'
gallíra	'collar'
zokni	'sock'
fusekla *	'sock'
harišňa	'stocking'
topánka *	'shoe'
čuka	'pike; shoe'
šusteri	'shoemaker'
kalapa	'hat'
šipka	'cap'
síja	'belt; strap'
kesťú(va)	'glove'
fáťuli	'veil'
gombo	'button'
gembešťú	'pin'
díso	'ornament, adornment'
kerpelece	'bracelet'
lánco	'chain; necklace'
ďenďo	'pearl; bead'
fingó(va)	'earring'
čelenka *	'headband'
tetováláši	'tattoo'
žepkendó	'handkerchief'
ronďo	'rag'
teríkezó	'towel'
kefe	'brush'

konťo	'plait, braid'
beretva	'razor'
kenéče	'ointment'
ďíkeri	'mirror'
guňhó(va)	'hut, shanty, hovel'
šátori	'large tent used for celebrations etc.'
udvara	'yard, court'
koňha	'kitchen'
soba †	'room (esp. inhabited by people)'
hejšígo	'room'
kaputa	'gate'
ajtóragastó	'doorpost'
záro	'lock'
reteza	'latch, door-bolt'
kallantú	'(a kind of) latch'
lakato	'padlock'
bloka	'window'
paló	'floor'
emeleto	'floor, storey'
falo	'wall'
kandalló	'(decorative) fireplace in a room'
kájha	'stove'
šporhelto *	'stove, (kitchen-)range'
kímíňi	'chimney'
létra	'ladder'
pokróca	'blanket'
takaró	'blanket'
paplaňi	'quilt, duvet'
larisa *	'(a kind of) blanket'
séko	'chair'
lampa *	'lamp'
lampáši	'lantern, lamppost, streetlamp, standard'
villalampa	'torch'

ďerťa	'candle'
póco	'shelf'
regáli *	'shelf with compartments'
serha	'roof'
gerenda	'beam'
oslopo	'post or pole'
deska	'board'
kémíveši	'mason, bricklayer'
tégla	'brick'
válka *	'adobe'
tábori *	'camp '
hinta	'swing'
hallóhinta	'hammock'
gazda *†	'farmer; householder, goodman; boss'
kerítíši	'fence'
sántinen	'to plough/plow'
kapálinen	'to dig'
ášó	'spade'
lapáta	'shovel'
kapa	'hoe'
gereble	'rake'
mago	'seed; grain; stone (of a fruit)'
kasálinen	'to mow'
šalló(va)	'sickle'
kasa	'scythe'
aratáši	'harvest'
gabona	'grain, corn (barley, oats etc.)'
semo	'grain, corn, kernel'
búza	'wheat'
arpa	'barley'
rožo †	'rye'
zabo	'oats'
riža *†	'rice'

mezó(va)	'grass'
sína	'hay'
ítetinen	'to plant'
tevo	'root'
ága	'branch'
fentó(va)	'loppings'
levele	'leaf'
virága	'flower'
telďo	'oak'
biko	'beech'
makko	'acorn'
cigarettázinen	'to smoke cigarettes'
pípa	'pipe'
tékó(va)	'tree stump'
terčo	'tree stem, tree trunk'
eňva	'sap (from a tree)'
citroňi	'lemon'
naranči	'orange'
banáno *	'banana'
bambusi *	'bamboo'
cukornádo	'sugar cane'
čóváňi	'nettle'
dógozinen	'to work'
hajtinen	'to fold'
ňírinen	'to cut with scissors or a similar instrument'
alló(va)	'scissors or shears'
dergelinen	'to rub'
akastinen	'to hang up'
ňominen	'to push; press; squeeze'
ňomkodinen	'to squeeze'
ňúzinen	'to flay; squeeze (fruits)'
šeprinen	'to sweep'
šeprú(va)	'broom'

sersámo	'manual workers' tool'
áčo	'carpenter'
ípítinen	'to build'
fúrinen	'to bore'
víšinen	'to hollow out'
fírísi	'saw'
kalapáči	'hammer'
ragastó	'glue'
kováči *	'blacksmith'
ílló	'anvil'
araňo	'gold'
ezišto	'silver'
ólmo	'lead'
plého	'tinplate'
aďago	'clay'
košara	'basket'
séňego	'carpet; mat; rug'
cekkeri	'netbag'
leďezó	'fan'
leďezinen	'to fan'
faraginen	'to carve'
sobrási	'sculptor'
sobro	'statue'
víšó	'chisel'
feštíko	'paint; make up'
íko	'chock, peg'
čipesi	'clothes-peg, clothes-pin'
éllešító	'sharpener', esp. 'whetstone'
mozdítinen	'to move [sth.]'
mozgatinen	'to move [sth.] back and forth'
mozginen	'to move'
čomagolinen	'to pack, wrap'
šodrinen	'to spin, twine, twist; roll (dough)'

gurítinen	'to roll, wheel, bowl'
čavarinen	'to twist; screw'
emelinen	'to lift; raise; pick up'
čepeginen	'to drip'
bukinen	'to sink; plunge, dive'
úsinen	'to swim'
freččelinen	'to splash, squirt'
repilinen	'to fly'
tǐúsinen	'to crawl; slide, slip'
másinen	'to climb; crawl'
terbekelinen	'to kneel'
gugolinen	'to crouch'
šétálinen	'to walk, take a walk, go for a walk'
tǐlinen	'to disappear'
ňomozinen	'to trace, trail; prospect for; enquire; pursue'
lešinen	'to watch for, lurk for, spy upon; prospect for'
vezetinen	'to drive; lead, rule, control'
definen	'to toss; poke at, thrust in, push into [so.]'
tolinen	'to push, to jostle; to wheel'
orságúto	'road'
hída	'bridge'
koči	'carriage, wagon, cart'
kereko	'wheel'
tengó(va)	'axle'
sánkó(va)	'sledge/sled'
hajó(va)	'ship'
čónako	'boat'
ladiko	'boat'
kompo	'ferry, raft, scow, pram'
evedzó	'oar, paddle'
evedzinen	'to row, paddle'
hajókórmáňi	'rudder'
kórmáňi	'rudder; steering wheel'

mačka (2)	'anchor'
mentinen	'to save, rescue'
šemmiššítinen	'to destroy'
kódúši	'beggar'
žugorotno	'stingy, avaricious; greedy'
kéčeno	'loan; debt'
sámla	'bill'
sázollíko	'percent; taxes'
fizetíši	'payment; salary, wages'
órabíro	'hourly wages'
kereškedinen	'to trade, merchandise; deal with; do business'
kofázinen	'to work at the market; barter; traffic, profiteer'
šeftelinen	'to trade, traffic'
kereškedó	'merchant'
kofa	'market woman, market person, stallkeeper; trafficker, chafferer'
pijarci	'market'
bóta	'shop, store'
izleto	'shop, store'
írtíko	'value; price'
óčóno	'cheap'
maradíko	'rest, remains'
ostinen	'to separate; apportion; divide'
alačonno	'low'
špicco	'top, peak; tipsiness'
feneke	'bottom'
vígo	'end'
heďešno	'mountainous; pointed'
sílo	'edge'
kezepo	'middle, centre'
jobno	'right'
balogno	'left'
balgačno	'left; left-handed'
kezé	'near'

mírinen	'to measure; weight'
síkno	'narrow'
mílno	'deep'
čekílno	'shallow'
lapošno	'flat'
igeňešno	'straight, direct'
horga	'fishhook, fishing rod, angle; hook'
keresto	'cross'
kocka	'die; cube; square; check'
gembelígno	'round'
kero	'circle'
guló	'ball'
číko	'strip; line; accent, háček (diacritic sign)'
luka	'hole, slot'
hašollóno	'similar'
váltostatinen	'to change [sth.]'
változinen	'to change'
nulla	'zero'
ezeri	'thousand'
čepo	'few, little; a few, a little (bit)'
šoro	'row; queue; turn; crowd'
čak	'only'
éšéno	'first'
utóšóno	'last'
páro	'pair, couple'
korá(n)	'early, early in the morning'
hajnábo	'at dawn, early in the morning'
kíšén	'late'
minďár	'right away, presently, immediately'
rekten	'right away, in no time, immediately'
folva	'right away, in no time, immediately'
ďoršan	'fast'
šijetinen	'to hurry'

kíšlinen	'to be late'
kezdinen	'to begin'
kezdíši	'beginning'
vígzinen	'to finish'
kísno	'ready'
mindig	'always'
furt *	'always, all the time; still'
álondóvan	'all the time, incessantly, unceasingly'
ďakran	'often'
valamikor	'sometimes; some time ago, in the old times'
níha	'sometimes, at times, now and then'
šoká	'long, for a long time'
šoha	'never'
újra	'again, anew'
újbú	'again, anew'
veradáši	'dawn'
hajnalo	'dawn, daybreak'
dílo	'midday, noon'
óra	'clock, watch; o'clock; hour'
hetfeno	'Monday'
keddo	'Tuesday'
serda	'Wednesday'
čiterteko	'Thursday'
pinteko	'Friday'
tavasi	'spring'
éso	'autumn/fall'
sagulinen	'to scent, sniff'
sagošno	'having a particular (neutral or good) smell'
izlinen	'to taste, have a particular taste'
hango	'sound; voice'
lárma	'noise'
hangošno	'loud, noisy'
čendešno	'quiet'

világítinen	'to give light; shine'
raďoginen	'to shine, glare, glitter'
síňo	'colour'
világošno	'light; bright; clear, obvious'
šetítno	'dark'
kíkno	'blue'
čípinen	'to pinch; sting; be hot (about food)'
durvavo	'rough, tough'
šímavo	'smooth'
élešno	'sharp'
tompavo	'blunt'
keňňíno	'light (in weight); easy'
nedvešno	'wet'
foróno	'hot, boiling hot'
ráncošno	'wrinkled'
lelko	'soul, spirit'
čudákozinen	'to be surprised, wonder, marvel'
bámulinen	'to gape, goggle; marvel'
ňalábolinen	'to embrace'
elelinen	'to embrace'
somorúšágo	'sadness, grief, gloom'
gondo	'concern, worry, anxiety'
šajnálinen	'to regret, be sorry; pity'
šajnálato	'regret, pity'
jajgatinen	'to wail, moan, groan'
idegeššígo	'nervosity; anger'
íriččígo	'envy'
kíňešno	'proud; delicate, squeamish'
bátorno	'brave, courageous; bold'
válostinen	'to choose, select; elect'
hazudinen	'to lie'
bečapáši	'deceit, fraud, bluff'
bočájtinen	'to forgive'

hiba	'mistake, error; defect; fault; blame'
íhéhótlanno	'insatiable, sateless, greedy (concerning food)'
gondúkozinen	'to think, reflect'
gondolinen	'to think, be of opinion'
találinen	'to hit the mark, strike home, nick; guess'
utánozinen	'to imitate'
gondulato	'thought; idea'
okošno	'wise, prudent'
butavo	'stupid, silly'
bambavo	'dumb, dull, booby, simple-minded'
tanuló	'pupil'
dijáko	'pupil, student'
iškoláši	'school-age child, pupil'
tanító	'teacher'
iškola	'school'
emlékezinen	'to remember, be able to retrieve from one's memory'
titkošno	'secret'
bistošno	'certain, sure, safe, dependable'
maďarázinen	'to explain'
sándíko	'intention'
kítelkedíši	'dubitation, scepticism; doubt'
kítelkedinen	'to doubt; suspect'
sikšígo	'need'
próbálinen	'to try; try out, test; practice'
próbinen	'to try; try out, test; practice'
mer(t)	'because'
mivel	'since, because, as'
vaď	'or'
há(t)	'well; yes'
šušoginen	'to whisper'
morginen	'to mumble'
fiťelinen	'to whistle'
ordítinen	'to yell, squeal, shriek, scream, shout'

šikójtinen	'to scream, shriek, yell'
ňerítinen	'to neigh, whinny; shriek'
hebeginen	'to stutter, stammer'
halgatinen	'to be silent'
felelinen	'to answer'
tagadinen	'to deny'
tiltinen	'to forbid'
nevezinen	'to call, term; name; designate'
jelentinen	'to announce, report; mean, have the meaning'
írinen	'to write'
papíri *	'paper; document; driver's licence'
tollo	'pen'
keňvo	'book'
veršíró	'poet'
šípa	'woodwind instrument'
dobo	'drum'
trombita	'trumpet'
čergó(va)	'rattle'
čerginen	'to rattle'
álmo	'state, country'
hazájo	'native country'
határi	'border, boundary, frontier'
nípo	'people; nation'
fajta	'clan'
vajda *	'Romani chieftain'
fé	'leader, boss, chief'
vezetó	'leader; driver'
bota	'walking stick, staff'
királi	'king'
rabo	'slave; prisoner'
solga	'servant'
sabadítinen	'to liberate, set free'
parančolinen	'to command, order'

engedinen	'to allow, permit, give way; let go; concede, submit'
baráto	'friend'
haveri	'buddy, pal, mate'
pajtáši	'buddy, pal, mate'
ellenšígo	'enemy'
somsído	'neighbour'
idegenno	'stranger; foreigner'
vendígo	'guest'
vendígelinav	'to host, regale'
šegítinen	'to help'
vídinen	'to defend; protect; prevent'
sokáši	'custom, habit, manner'
feleškedíši	'quarrel'
talákozinen	'to meet'
marakodinen	'to quarrel, row, brawl, wrangle; fight, scarp'
háborúzinen	'to lead war, war'
háború	'war; battle'
bíkeššígo	'quietude, serenity; peace'
feďveri	'weapon'
itleko	'weapon'
pariťťa	'sling'
ňila	'bow and arrows'
kardo	'sword; saber'
puška *	'gun, rifle'
šišako	'helmet'
torňo	'tower'
ňeríši	'victory; gain'
vestíši	'defeat, loss (in a game etc.)'
támodáši	'attack'
fogla	'captive; prisoner'
bertenéreši	'prisoner'
éršígo	'guard'
viďázó	'watchman, keeper, guard'

halási	'fisherman'
žinego	'line; fishing line'
hálló	'fishnet'
čaló, čálló	'bait'
vadásinen	'to hunt'
levinen	'to shoot'
lédezinen	'to shoot (frequentative)'
múlinen	'to go by, elapse; cease; miss'
čabda	'trap, catch, pitfall'
tervíňa	'law'
bíróšágo	'court of justice; judgement'
bíró	'judge; referee'
tanú	'witness'
ítílinen	'to condemn'
bínešno	'guilty, sinful'
hibášno	'false; guilty'
bíntelenno	'innocent'
bíntetíši	'penalty, punishment'
berteno	'prison, jail'
ďilkoššágo	'murder'
erésakolláši	'rape'
hito	'belief, faith; denomination, religion'
oltári *	'altar'
sentno	'holy, sacred, saint'
prédikálinen	'to preach'
áldinen	'to bless'
átkozinen	'to curse'
bétlinen	'to fast'
poklo	'hell'
sellemo	'spirit, ghost'
jelentíši	'announcement, report; meaning, something meaningful; omen'
rádió	'radio'
televízijó	'television; TV set'

telefono	'telephone'
bicigli	'bicycle'
motorka *	'motorcycle'
avutó	'car'
busi	'bus'
vonato	'train'
repilló	'airplane'
villaňi	'electricity, electric light'
ellemo	'battery'
baterka *	'battery'
fékezinen	'to brake'
motori *	'motor, engine'
gípo	'machine'
nafta *	'petroleum'
petrolímo *	'petroleum'
korházo	'hospital'
sestrička *	'nurse'
tabletta	'pill, tablet'
inekció	'injection'
semivego	'spectacles, glasses'
ministeri *	'minister'
čendéršígo	'police'
čendéri	'policeman'
voďičáko *	'driver's license'
tábla	'plate; blackboard; license plate'
kerestlevelo	'birth certificate'
válostáši	'choice, selection; election'
címo	'address'
sámo	'number'
ucca	'street'
póšta	'post, mail; post office'
billego	'postage stamp; seal'
kárťa	'card; letter'

kípešlapo	'postcard'
banka *	'bank'
čapo	'tap, faucet'
moždó	'washbasin, sink'
véce	'toilet'
klozeto	'toilet'
madraci	'mattress'
bátogó	'tin, can'
konzerva *	'tin, can (canned goods)'
čavaró	'screw'
šrófo	'screw'
čavarhúzó	'screwdriver'
míjaňago	'plastic'
bomba	'bomb'
míhele	'workshop'
cigaretta	'cigarette'
újšágo	'newspaper'
naptári	'calendar'
mozi	'cinema, movies; film, movie'
zene	'music'
teja	'tea'
kávéja	'coffee'

Slovak and/or Czech

* must be Slovak; † must be Czech

miminko †	'baby'
žraloko	'shark'
velriba	'whale'
úhori *	'freshwater eel'
škorpióni	'scorpion'
kengura *	'kangaroo'
búvoli †	'buffalo'
krokodíli	'crocodile'

paprečko *	'toe, claw'
boka (2)	'hip (external)'
pepšo †	'black pepper'
špendlíko	'pin'
šperko	'jewel'
stano	'tent (tourist)'
futro †	'doorpost'
zárubňa	'doorpost'
polička	'shelf (decorative)'
klenba	'arch'
malta	'mortar (building material)'
jezeďáko †	'cooperative farmer'
palma	'palm tree'
kokosi	'coconut'
šiška	'cone'
cíno	'tin'
bumerango	'boomerang'
účeto	'bill'
daňe	'taxes'
víkhodo	'East (= eastern Slovakia)'
západo	'West (= western Slovakia)'
šípo	'arrow'
ošťepo	'spear'
helma	'helmet'
baxo †	'jailer, jail guard'
súdo *	'court'
víla	'fairy'
škráteko *	'elf'
vláda	'government'
prezidento	'president'

Vlax Romani

žuvárno	'stingy'

krísa 'Vlax community-internal court'